Models for Enzyme Kinetics, Including Cellulases and the Enzymatic Degradation of Biomass

by John William Dingee

This thesis/dissertation document has been electronically approved by the following individuals:

Anton,Alan Brad (Chairperson)

Walker,Larry P (Minor Member)

Wilson,David B (Minor Member)

MODELS FOR ENZYME KINETICS, INCLUDING CELLULASES AND THE ENZYMATIC DEGRADATION OF BIOMASS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

John William Dingee

August 2010

© 2010 John William Dingee

MODELS FOR ENZYME KINETICS, INCLUDING CELLULASES AND THE ENZYMATIC DEGRADATION OF BIOMASS

John William Dingee, Ph.D.

Cornell University 2010

This dissertation contains several enzyme kinetics studies, each demonstrating the importance of dimensional analysis in justifying rate expressions and the usefulness of perturbation techniques in generating approximate time-dependent solutions. Our focus has been on the action of cellulases, enzymes which hydrolyze the glycosidic bonds of cellulose, releasing sugars which can be used as feedstock for renewable commodities and fuels. The process involves two fundamental steps at the solid-liquid phase boundary: enzyme binding and hydrolysis. The development of detailed mathematical models plays an essential role in understanding the mechanism of this heterogeneous process.

First, we present two homogenous enzyme kinetics studies, where reactions are carried out in a single well-mixed liquid-phase. The first study focuses on the Michaelis-Menten (MM) mechanism for enzyme kinetics. The mass balances for MM kinetics are simple to write but deceptively difficult to solve; consequently, various approximate solutions have been offered over the past 50 years, none of which works in all cases. We generate a uniformly-valid time-dependent solution that converges accurately for any combination of initial conditions, and we systematically define where each previous solution fits within the new solution. In the second study, we investigate the hydrolysis of p-nitrophenyl cellobioside by *Thermobifida fusca* Cel5A, a retaining endocellulase, and develop a mechanistic model for its hydrolysis and transglycosylation in the enzyme's active site. Our work extends previous treatments by providing criteria that justify the use of the quasi-steady-state approximation

(QSSA) and provide an integrated form of the resulting rate expression. The results can be extended to other retaining glycosyl hydrolases acting on nitrophenol glycosides.

Next, we present a heterogeneous two-phase kinetics model for the hydrolysis of dense cellulose fibers by cellulase enzymes. The model reveals that the shapes of the time-dependent liquid phase concentration curves can appear nearly identical even when different underlying mechanisms are dominant. The results substantiate the importance of including penetration and local reaction history in cellulase kinetics models and of more carefully measuring the evolving shapes of fibers when applying such models.

Finally, we review our collaborative efforts to develop a synthetic cellulose substrate with controlled microstructure and physical properties.

BIOGRAPHICAL SKETCH

John (Jack) Dingee was born September 15, 1981 in Syracuse, New York and grew up in the Schoharie river valley near Albany, New York, with his mother, father, and five siblings. As a young man, John enjoyed skateboarding, wrestling, and hunting, and held a paper route for many years in the neighboring village. He also met and dated his future wife, Donna-Rose, who rode the same #3 school bus. After graduating from Duanesburg Central High School in June of 1999, John attended Hudson Valley Community College and then Rensselaer Polytechnic Institute where he enrolled in the Mechanical Engineering program. He changed his major to Chemical Engineering after a long talk with Michael Abbott, and earned his Bachelor's Degree in May of 2003. In August of the same year, John and Donna-Rose were married and moved to Ithaca where John began his studies at Cornell. After graduation, John plans to seek an industrial research position or an academic career teaching at a small University.

To my wife Donna-Rose

ACKNOWLEDGMENTS

I'd like to thank my committee members for their involvement and support. This degree would not have been possible without my advisor, Brad Anton, whose honest and thoughtful professional and personal advice were invaluable in the process. Brad has taught me a lot about modeling, analysis, and technical writing. My minor advisor, David Wilson, generously allowed me to work and sit in his lab for four years, purifying and studying cellulases with his students. He has provided me with extensive feedback throughout my studies that have deepened my understanding of cellulase enzymes and glycosyl hydrolases. I'd like to thank my department co-chair Larry Walker, who was instrumental in my receiving an MGET fellowship and introduced me to cellulases in his course on biomass conversion. I'd also like to thank Yong Joo and his student, Heidi Park, for their collaborative work with me involving electrospun cellulose. From the Wilson Lab, I'd like to personally thank Diana Irwin, Thu Vuong, Yongchao Li, and Colleen McCrath, with whom I could discuss and debate the fine details of my experiments and cellulase enzymes. I'd also like to thank to Paul Steen for acting as proxy for Larry Walker at my thesis defense. I would like to thank my family for visiting often and supporting me. Most of all, I thank my wife Donna-Rose for her untiring love and support throughout my graduate career.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

GH - Glycosyl Hydrolase

CD - Catalytic Domain

CBM - Catalytic Binding Module

LR – Linker Region

CrI - Crystallinity Index

WAXS - Wide Angle X-ray Spectroscopy

DP - Degree of Polymerization

BMCC - Bacterial Microcrystalline Cellulose

FP - Filter Paper

PASC - Phosphoric Acid Swollen Cellulose

CMC - Carboxymethyl-Cellulose

KMC - Kinetic Monte-Carlo

EHF - Easily Hydrolysable Fraction

ESC - Electrospun Cellulose

MM - Michaelis-Menten

$O(x)$ - "Order of x" in an Asymptotic Sense

ODE – Ordinary Differential Equation

PDE – Partial Differential Equation

QSSA - Quasi-Steady-State Approximation

sQSSA - Standard Quasi-Steady-State Approximation

tQSSA - Total Quasi-Steady-State Approximation

rQSSA - Reverse Quasi-Steady-State Approximation

UVA - Uniformly-Valid Approximation

HTA - Heineken, Tsuchiya, and Aris

SM - Schnell and Mendoza

TE - Tzafriri and Edelman

Cel5Acd - *Thermobifida fusca* Cel5A Catalytic Domain

G1 - Glucose

G2 - Cellobiose

G3 - Cellotriose

G4 - Cellotetraose

pNP - p-nitrophenol

pNP-G1 - para-nitrophenol-Glycoside

pNP-G2 - para-nitrophenol-β-D-Cellobioside

pNP-G4 - p-nitrophenol Cellotetraose

NaAc - Sodium Acetate

NaAz - Sodium Azide

TLC - Thin Layer Chromatography

DI Water - De-ionized Water

NMMO - N-methylmorpholine-N-oxide

SEM - Scanning Electron Microscopy

AFM - Atomic Force Microscopy

A280 - Absorbance at 280 nm

PAHBAH – Para-Hydroxybenzoic Acid

LIST OF SYMBOLS

CHAPTER 1

E - Free Enzyme

σ - Binding Site

B - Bound Enzyme

P - Sugar Product

α - Product Freed per Binding Site

CHAPTER 2

E - Enzyme Concentration

S - Substrate Concentration

C - Enzyme-Substrate Complex Concentration

P - Product Concentration

E_T - Initial Enzyme Concentration

S_T - Initial Substrate Concentration

k_1, k_{-1}, k_2 - Elementary Rate Constants

K - Van Slyke-Cullen Constant

K_D - Dissociation Equilibrium Constant

K_M - Michaelis-Menten Constant

$V_{\max}$ - Maximum Reaction Rate

V_0 - Initial Reaction Rate

β - Maximum Complex Concentration (Leading Order Estimate)

θ_C - Characteristic Timescale for Complex Formation

θ_P - Characteristic Timescale for Product Formation

t - Time

τ - Time Scaled Relative to the Characteristic Time for Complex Formation

T - Time Scaled Relative to the Characteristic Time for Product Formation

x, X - Dimensionless Complex Concentration (Inner Problem, Outer Problem)

y, Y - Dimensionless Product Concentration (Inner Problem, Outer Problem)

z, Z - Dimensionless Substrate Concentration (Inner Problem, Outer Problem)

w, W - Dimensionless Enzyme Concentration (Inner Problem, Outer Problem)

γ - Maximum Fraction of Initial Enzyme in Complex (Leading Order Estimate)

μ - Maximum Fraction of Initial Substrate in Complex (Leading Order Estimate)

η, ϕ - Kinetic Branching Ratios

δ - Regular Perturbation Parameter

ε - Singular Perturbation Parameter

κ, λ - Parameters of Related Model (Defined In Text)

x_0, X_0 - Leading Order Solution for x, X

x_δ, X_δ - $O(\delta)$ Solution for x, X

$x_\varepsilon, X_\varepsilon$ - $O(\varepsilon)$ Solution for x, X

y_0, Y_0 - Leading Order Solution for y, Y

y_δ, Y_δ - $O(\delta)$ Solution for y, Y

$y_\varepsilon, Y_\varepsilon$ - $O(\varepsilon)$ Solution for y, Y

x_{match} - Matching Solution for x

y_{match} - Matching Solution for y

A, B, C - Integration Constants

CHAPTER 3

E_T - Initial Enzyme Concentration

S_T - Initial Substrate Concentration

E - Enzyme (Cel5Acd)

S - pNP-G2 Substrate

C_1 - Michaelis Complex

C_2 - α-Glycosyl/Enzyme Intermediate

C_3 - Second Michaelis Complex

R - pNP Reporter

R^* - Secondary pNP-G1 Reporter

P - pNP Product

P^* - Secondary G3 product

$k_1, k_{-1}, k_2, k_3, k_4, k_{-4}, k_5$ - Elementary Rate Constants

K_m, K_m^* - Pseudo-Michaelis-Menten Constants

$\widetilde{k}_{cat}, \widetilde{K}_m, \widetilde{K}_{m,2}, \widetilde{k}_{cat,2}$ - Lumped Kinetics Parameters

$\overline{k}_{cat}, \overline{K}_m$ - Apparent Michaelis-Menten Constants

R_∞ - Final Reporter Concentration

f - Fraction of pNP in Initial Substrate Converted to R

ω - Transglycosylation Parameter

χ^2 - Best-Fit Parameter

K - Lumped Model Parameter

β_i - Scaling Factor for Complex, C_i, where i = 1, 2, 3

X_i - Dimensionless Complex Concentration, where i = 1, 2, 3

Z - Dimensionless Substrate Concentration

Y - Dimensionless Reporter Concentration

t - Time

t_R - Characteristic Timescale for Reporter Formation

T - Time Scaled Relative to the Characteristic Time for Reporter Formation

γ_i - Maximum Fraction of Enzyme in Complex, where i = 1, 2, 3

δ_i - Lumped Model Parameter, where i = 1, 2, 3, 4

ε_i - Timescale Ratios, where i = 1, 2, 3, 4

c_i - Lumped Model Parameter, where i = 1, 2, 3, 4, 5

CHAPTER 4

E_0 - Initial Liquid Phase Enzyme Concentration

M_0 - Initial Liquid Phase Cellulose Concentration

σ_0 - Initial Point-wise Binding Site Concentration

R - Initial Fiber Radius

ρ_c - Initial Cellulose Density

$\hat{r}$ - Radial Position of Accessibility Boundary

v - Speed of Accessibility Boundary

ψ - Initial Solid Volume Fraction of Cellulose

k_1, k_{-1}, k_2 - Elementary Reaction Rate Constants

K_D - Dissociation Constant

α - Ratio of Product Available per Binding Site

$R_B(r,t)$ - Point-wise Binding Rate

$R_R(r,t)$ - Point-wise Release Rate

$R_H(r,t)$ - Point-wise Hydrolysis Rate.

θ_b - Characteristic Timescale for Binding

θ_R - Characteristic Timescale for Penetration

$\theta_{\bar{\sigma}}$ - Characteristic Timescale for Hydrolysis

θ_D - Characteristic Timescale for Diffusion

L - Characteristic Path Length for Diffusion

ϕ - Thiele Modulus

ε - Ratio of Characteristic Timescales for Binding-to-Hydrolysis

δ - Ratio of Characteristic Timescales for Penetration-to-Hydrolysis

t - Time

t' - Local Reaction Time

T - Time Scaled Relative to the Hydrolysis Timescale

T' - Local Reaction Time Scaled Relative to the Hydrolysis Timescale

r - Radial Position

ρ - Dimensionless Radial Position

$\hat{\rho}$ - Dimensionless Radial Position of Accessibility Boundary

$E(t)$ - Mixing-Cup Enzyme Concentration

$B(t)$ - Mixing-Cup Bound Enzyme Concentration

$P(t)$ - Mixing-Cup Product Concentration

$E(r,t)$- Enzyme Concentration in Reacting Zone

$P(r,t)$- Product Concentration in Reacting Zone

$b(r,t)$ - Bound Enzyme Concentration in Reacting Zone

$\sigma(r,t)$ - Binding Site Concentration in Reacting Zone

$\overline{\sigma}(r,t)$- Total Binding Site Concentration in Reacting Zone

$A(r,t)$ - Accessible Area for Enzyme Binding in Reacting Zone

$D(r,t)$ - Effective Diffusivity of Enzyme in Reacting Zone

x - Dimensionless b

z - Dimensionless σ

$\bar{z}$ - Dimensionless $\overline{\sigma}$

X - Dimensionless B

W - Dimensionless E

Y - Dimensionless P

Z - Dimensionless Mixing-Cup Binding Site Concentration

$\overline{Z}$ - Dimensionless Mixing-Cup Total Binding Site Concentration

β - Maximum Point-wise Bound Enzyme Concentration (Leading Order Estimate)

X - Dimensionless Liquid Phase Bound Enzyme Concentration

μ - Maximum Factional Amount of Occupied Binding Sites

ξ - Maximum Fractional Amount of Enzyme Bound from the Liquid Phase

x_0, x_1 - $O(1)$, $O(\xi)$ Solution for x

$\bar{z}_0$, $\bar{z}_1$ - $O(1)$, $O(\xi)$ Solution for $\bar{z}$

X_0, X_1 - $O(1)$, $O(\xi)$ Solution for X

$\overline{Z}_0$, $\overline{Z}_1$ - $O(1)$, $O(\xi)$ Solution for $\overline{Z}$

λ_1, λ_2 - Eigenvalues

ω - Model Parameter

$t_{1/2}$ - Reaction Time for Half of Total Product Formation

$B_{\max}$ - Maximum Mixing-Cup Bound Enzyme Concentration

$P_{\max}$ - Maximum Mixing-Cup Product Concentration

C_i - Integration Constant, where i = 1 through 4.

a_i - Lumped Model Parameters, where i = 1 through 7

A_i - Lumped Model Parameters, where i = 1 through 28

B_i - Lumped Model Parameters, where i = 1 through 6

CHAPTER 1

THE ENZYMATIC HYDROLYSIS OF CELLULOSE

1.1 Introduction

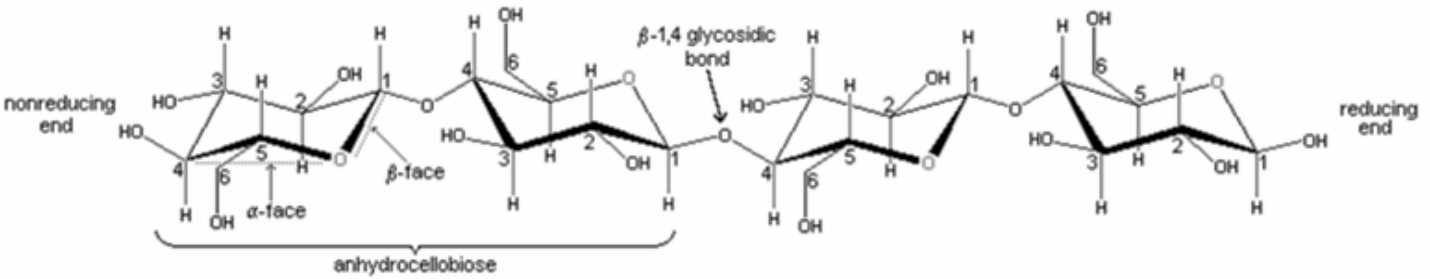
Cellulose is a long-chain, unbranched homopolymer that is the major component of plant biomass. The intra-chain β-1,4 linked glycosidic bonds of cellulose are hydrolyzed by a class of modular enzymes known as cellulases, releasing glucose and soluble oligomers for use as feedstock in the production of renewable commodities and fuels **[1]**. The enzymatic processing of cellulose has long been envisioned as a "clean" alternative to petroleum, offering a biologically renewable catalyst and feedstock that can be combined at mild operating conditions with high bond cleavage specificity. The large-scale commercial utilization of cellulose is currently limited by the recalcitrance of biomass, which leads to high pretreatment costs, long enzyme exposure times, and low yields **[2]**.

1.2 Cellulose, Lignocellulose & Cellulosic Substrates.

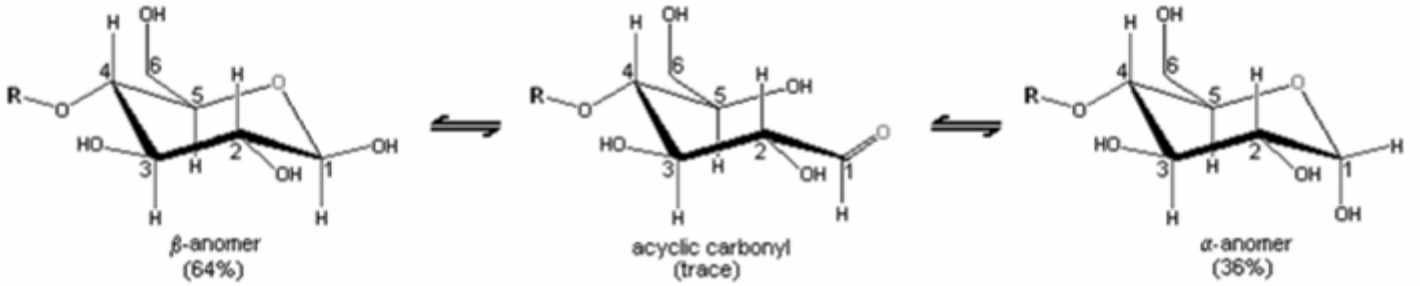
Cellulose. Cellulose is an unbranched β-1,4 linked homopolymer of anhydroglucose (D-glucose), with a natural degree of polymerization ranging from ~10^2 to ~10^5 monomer units **[3, 4]**. In the polymer molecule, the sequential glucopyranose residues are in their lowest energy $^4C_1$ chair conformation and rotated sequentially by 180^o with respect to the ring plane, making the repeat unit of cellulose anhydrocellobiose. Cellulose has polarity with a reducing and non-reducing end. While the non-reducing end contains a chemically stable hydroxyl group at C4, the

Figure 1.1: Basic Cellulose Chemistry. Cellulose is a repeating polymer of anhydrocellobiose (~10 Å long) with a reducing and nonreducing end (cellotetraose is shown in the top figure). Each D-glucopyranose unit (~5 Å long) is in a $^4C_1$ chair conformation where the C4 carbon is above the plane of the C1 carbon. The glycosidic bond is formed between the C1 and C4 carbons at the β-faces of two residues, hence a β,1-4 glycosidic bond. The glycosidic bonds, the C6 carbon, and C2 and C3 hydroxyl groups extend out in the plane of the ring. At the reducing end, the C1 hydroxyl can undergo mutarotation between two anomers via an acyclic carbonyl intermediate (a reducing end is depicted in the bottom figure). The natural concentrations of each anomer are given for glucose and cellobiose.

Cellotetraose Structure



Reducing End Chemistry



reducing end contains a hemi-acetal group and undergoes continuous mutarotation between two anomeric states via an acyclic carbonyl intermediate at C1, the anomeric carbon. Mutarotation rates for the reducing end are likely similar to glucose and cellobiose, which are on the order of 10^{-4} s^{-1} at 27 ^{o}C **[5]**. The end is said to be reducing because the reactive aldehyde intermediate can be oxidized to carboxylic acid, reducing the complementary reacting molecule. The β-glycosidic bonds and the hydroxyl groups at C2, C3, and C6 are equatorial with respect to the glucopyranose rings. Intramolecular hydrogen bonding between the equatorial hydroxyl groups adds rigidity to the already straight-repeating-unbranched structure of cellulose. The face of the glucopyranose rings are hydrophobic and can participate in van der Waals bonding with the glucopyranose rings of adjacent chains **[6, 7]**. The axial hydroxyl groups can also participate in hydrogen bonding with adjacent cellulose chains. Due to these intermolecular bonding interactions, cellulose molecules longer than six monomer units are insoluble in water.

Insoluble cellulose chains are organized into a semi-ordered, heterogeneous phase. Sheets of stacked cellulose chains with defined hydrogen bonding patterns and van der Waals interactions form crystalline regions within the solid phase. These crystalline regions form various stable states or polymorphs **[8]**; the polymorph of native cellulose is type I, from which other polymorphic states can be achieved though chemical processing **[3]**. Cellulose type I can be further divided into a combination of two types, type Iα and type Iβ; the ratio of type Iα to Iβ is high in cellulose produced by bacteria and algae and very low in cellulose produced by higher plants **[9]**. After cellulose I, cellulose type II is the second most commonly studied type of cellulose, formed by regeneration (solublization followed by precipitation) or mercerization (alkali treatment). In comparison to the other polymorphs, including type I, type II is considered to be in the most favorable thermodynamic conformation. The non-

crystalline or amorphous regions of cellulose are also insoluble with extensive but random hydrogen bonding. Insoluble cellulose is a mixture of crystalline, amorphous, and so-called para-crystalline regions that share characteristics of both types. The crystallinity Index or CrI (%) is the fraction of solid in the crystalline state, as measured by Wide Angle X-ray Spectroscopy (WAXS).

Although the crystalline suprastructure of cellulose is fairly well established, the subsequent ultrastructure is not. From microscopic inspection of various forms of cellulose, a morphological hierarchy has been developed to describe its ultrastructure: elementary fibrils (~3 nm in diameter) are organized into microfibrils (~10-30 nm in diameter), which are in turn organized into macrofibrils (60-360 nm in diameter) **[10]**. The popular fringed fibrillar model for elementary fibrils suggests a single chain may pass through both crystalline regions (crystallites) and para-crystalline regions before terminating. A more detailed review of cellulose ultrastructure can be found elsewhere **[3, 8, 11]**.

Lignocellulose. Cellulose rarely occurs in its pure form in nature but is present in the plant cell wall (often as microfibrils), embedded with hemicellulose, lignin, pectin, and various cell wall proteins **[12]**, in a general class of materials referred to as lignocellulose. Hemicellulose is not cellulose but a general term for carbohydrate polymers in the cell wall that can be extracted with alkali treatment **[13]**: these polymers include xylan, glucuronoxylan, arabinoxylan, glucomannan, and xyloglucan. Lignin is a hydrophobic, randomly-linked, heterogeneous, aromatic, non-polysaccharide polymer. Pectin is a general class of complex-multifunctional polysaccharides that form a hydrated cross-linked network within plant cell walls. Polysaccharide structures within pectin include homogalacturonans and highly branched rhamnogalacturonans I and II **[14]**. Many excellent reviews about the plant cell wall can be found elsewhere **[15, 16]**. The aggregation of cellulose into an

insoluble, heterogeneous, hydrogen-bonded cell wall matrix presents a physical and chemical barrier for cellulase enzymes.

Model Substrates. To sidestep the complexity of naturally occurring lignocellulosics, cellulase researchers use widely available and relatively pure forms of cellulose known as model substrates **[10]**. There are many noteworthy model substrates. Valonia cellulose is straight, highly crystalline microfibril, about 25 nm in width, purified from the green sea algae *Valonia ventricosa* **[17, 18]**. Bacterial microcrystalline cellulose (BMCC) is a purified highly crystalline fiber of bundled type I microfibrils typically secreted from *Acetobacter xylinum*. Microcrystalline cellulose powders with intermediate crystallinity such as Avicel and Sigmacell are commercially available and are produced from finely ground wood pulp that has been treated with dilute acid. Whatman #1 filter paper (FP) is a commonly used paper substrate that is hole-punched into test tubes: one disc per tube, reacted for a fixed period of time, gives the so-called "FP activity". Phosphoric acid swollen cellulose (PASC) **[19]**, is made from swelling other model substrates such as Avicel in phosphoric acid, creating an almost completely amorphous phase. Carboxymethyl-cellulose (CMC) is a long-chain cellulose polymer where carboxymethyl groups have been substituted at the hydroxyl groups (usually around 15%), making the molecule soluble at much longer lengths than native cellulose. Other soluble cellulosic model substrates include soluble cellulo-oligosaccharides and their modified derivatives. The most common modified cellulo-oligosaccharides have chromophores or fluorophores attached at their reducing end.

1.3 Glycosyl Hydrolases & Cellulolytic Organisms

Cellulolytic organisms, which are predominantly fungi or bacteria, utilize

lignocellulose as their carbon source by producing an ensemble of glycosyl hydrolases: a class of enzyme that hydrolyze the glycosidic bonds of cell wall carbohydrates. Glycosyl hydrolases are named after their respective carbohydrate substrate and include cellulases, chitinases, mannanases, amylases, lactases, and xylanases **[20]**. An up-to-date database of these "Carbohydrate Active enZymes" is available on the CAZy website **[21]**.

Cellulases produced by these organisms come in two varieties: as complexed cellulosome systems and as non-complexed systems. The system produced by a microorganism appears to depend on its metabolic environment **[22]**. Cellulases produced by aerobic fungi such as *Trichoderma reesei* **[23, 24]** or *Humicola insolens* **[25]** and aerobic bacteria such as *Thermobifida fusca* **[26]** or *Cellulomonas fimi* **[27]** are typically secreted into the solution surrounding the microorganism and act as cell-independent functional enzymes in what are called free or non-complexed cellulase systems. Anaerobic bacteria, such as *Clostridium thermocellum* **[28]** or *Clostridium cellulolyticum* **[29, 30]**, typically found in soil or sediments, produce cellulase systems where the hydrolytic protein components aggregate into large multi-protein complexes or organelles, called cellulosomes **[31]**. In general, "cellulase" refers to the non-complexed systems while "cellulosome" refers to the complexed systems.

Cellulases are functionally categorized (how they degrade their substrate) as endocellulases, exocellulases, and β-glucosidases: Endocellulases have active site clefts and cleave within the interior of cellulose chains; exocellulases have active site tunnels and cleave glucose or cellobiose successively from the ends of cellulose chains; and β-glucosidases cleave the β-1,4 glycosidic bonds of soluble cello-oligosaccharides, especially cellobiose. Exocellulases, and some endocellulases **[32-34],** are processive - they can bind to and move along cellulose chains (in solution or on a surface) causing sequential cleavage of glycosidic bonds before an unbinding

event, with the direction of processivity being cellulose-polarity-dependent **[35]**. Thus, exocellulases are subdivided as either nonreducing-end specific or reducing-end specific. The combination of these three cellulases (endo/exo and β-glucosidase) has historically represented the prototypical model cellulase system required to efficiently hydrolyze crystalline cellulose. Cellulolytic organisms that can effectively utilize cellulose as a carbon source usually secrete either a complexed or non-complexed system of cellulases with these three functional activities.

1.4 Cellulase Protein Structure & Function

Cellulases are modular enzymes - they contain interconnected globular domains, each with distinct biochemical structures and functions. The most common modular configuration for cellulase enzymes is a catalytic domain (CD) connected to a carbohydrate binding module (CBM) via a flexible linker region (LR). This CD-LR-CBM or CBM-LR-CD arrangement is shared among glycosyl hydrolases. Cellulases are said to be modular because each domain (CD or CBM) can be isolated or cloned and the resulting polypeptide studied independently regarding its structure and function. Furthermore, the CDs and CBMs can be swapped between homologous proteins and maintain their respective polypeptide folds and activities.

Due to the domain modularity of most glycosyl hydrolases, their CDs and CBMs are separately grouped into families based on sequence similarities and hydrophobic cluster analysis **[36]** and then further into superfamilies or clans based on common folds and catalytic mechanisms **[37]**. Currently, glycosyl hydrolase domains are grouped into over 100 CD families within at least 14 different clans and over 40 CBM families within at least 7 different clans **[38]**. Of the over 100 GH CD families,

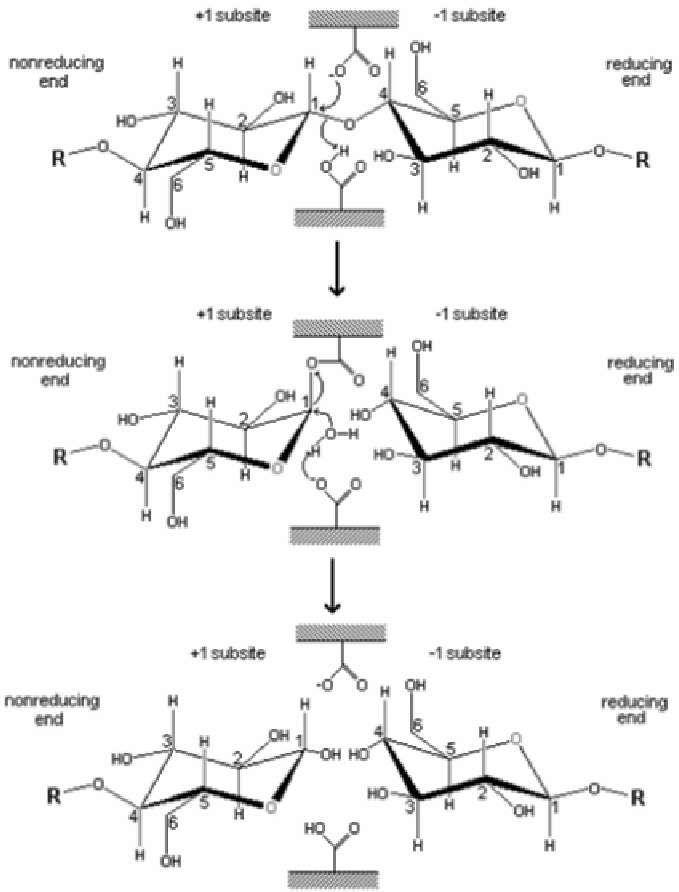
cellulases have been found in families 5, 6, 7, 8, 9, 12, 26, 44, 45, 60 and 61, with the most well studied families being 5, 6, 7, and 9.

Catalytic Domain. The CD is solely responsible for the hydrolytic action of cellulase enzymes. Cellulase CDs are typically globular folds with molecular weights ranging from 20 to 40 kDa and polypeptide lengths ranging from 150-450 residues **[39]**. Unlike the CDs of other enzymes, such as amylases, the CDs of cellulases do not share a common fold but display a wide vary of them **[40]**. Cellulolytic organisms produce GHs within multiple fold families. Some of the families are shared between bacteria and fungi, such as family 6, while family 7 are found only in fungi and others like family 48 are found only in bacteria. The endo/exo specificity is not conserved in fold families.

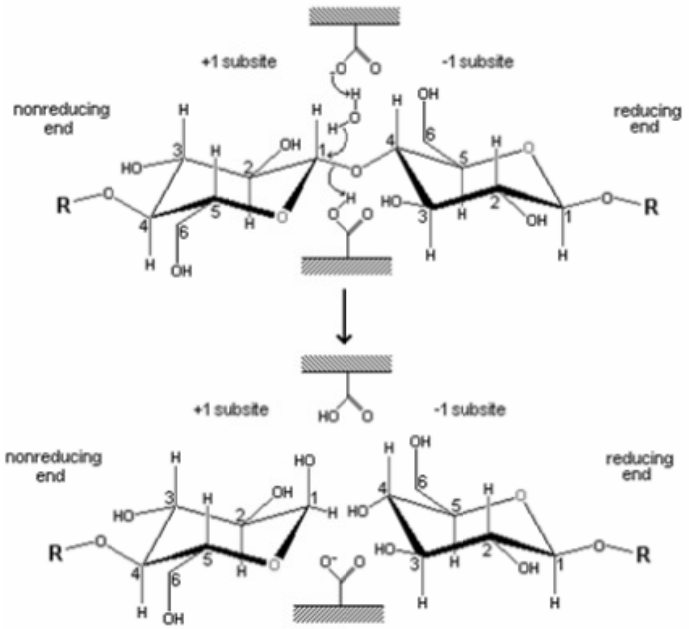
Cellulases bind cellulose with fixed polarity within their CD active site by forming hydrogen bonds with the substrate's equatorial hydroxyls and by hydrophobically stacking aromatic residues with the glucopyranose rings of the substrate, which are bound at specific subsites in that active site. Subsites are numbered from the reducing (negative #) to non-reducing (positive #) end of the substrate, with hydrolysis occurring at the glycosidic bond between the -1 and +1 subsites **[41]**. Cellulases cleave glycosidic bonds using general acid/base catalysis with either an inverting or retaining stereochemical mechanism **[42, 43]**, which is conserved among glycosyl hydrolase fold families. The general acid/base mechanism usually involves two opposing carboxyl residues within the active site, although exceptions can be found in family 6 **[39]** and family 4 **[44-46]**. In the inverting mechanism, a water molecule is charged by a catalytic acid before carrying out a nucleophilic attack at the C1 carbon, breaking the glycosidic bond. At the same time, the other catalytic acid donates a proton to the glycosidic oxygen leaving group. The single-step bond cleavage inverts the anomeric conformation at C1. In the retaining

Figure 1.2: Stereochemical Mechanisms for Glycosyl Hydrolases. The inverting mechanism results in an inversion of the anomeric configuration at the C1 carbon (right) and the retaining mechanism results in retention of the anomeric configuration at the C1 carbon (left).

Retaining



Inverting



mechanism, a catalytic residue carries out the nuclueophilic attack instead of water, forming a covalent glycosyl-enzyme bond. In a second step, a water molecule attacks and breaks that bond, freeing the carbohydrate product. The non-substrate-bonding catalytic acid acts as a generalized acid in the first reaction and donates a proton to the leaving group, and acts as a generalized base in the second reaction and removes a proton from the water nucleophile. The anomeric conformation at C1 is inverted twice, thus retaining its conformation. A critical aspect of the retain mechanism is that the attacking group in the second step can be any R-OH group that is sufficiently charged by the catalytic acid, including other carbohydrates, resulting in transglycosylation. The inverting or retaining stereochemistry of glycosyl hydrolases can be observed directly with ^{1}H-NMR **[47]**.

Carbohydrate Binding Module. Carbohydrate binding modules found in glycosyl hydrolases recognize and form stable non-covalent bonds with their carbohydrate substrates **[48]**. Type A CBMs found in cellulases, sometimes called cellulose binding domains (CBD), bind specifically to insoluble cellulose surfaces through a planar binding-face, utilizing hydrophobic and hydrogen bonding interactions **[38]**. The aromatic residues responsible for hydrophobic stacking are strongly conserved: usually two or three are spaced out along the binding face at about 10 Å allowing the residues to interact with every other glucopyranose ring **[49]**. Unlike CD fold families, CBM fold families are never shared among fungi and bacteria. Small, 30-40 residue, family 1 CBMs are found exclusively in fungi, whereas larger, ~100 residue, family 2 CBMs are found exclusively in bacteria. In spite of their difference in molecular weight, their binding faces have been shown to have similar size and function **[50]**.

The principle role of the CBM is to bind the entire protein at the phase boundary, increasing the propensity of the attached CD to encounter and hydrolyze

cellulose chains. Removing the CBM causes a substantial reduction in activity on highly crystalline cellulose, a less drastic reduction on amorphous cellulose, and does not change activity on soluble substrates **[39]**. Others have suggested a more active role for the CBM, where it forms a wedge and ploughs along the top layer of the crystalline cellulose, feeding the cellulose into the CD. Although isolated CBMs do not stimulate isolated CD activity **[51]** they have been shown to disrupt the cellulose structure **[52-54]**. The fact that many cellulolytic organisms secrete interlinked or free CBMs with no attached CDs, and secrete proteases that promote the freeing of CBMs from their full-length protein suggests that there is an additional role for CBMs in the hydrolysis process.

Linker Region. The CBM is covalently attached to the CD via a LR. The LR is typically 20-40 residues in length, heavily glycosylated, and rich in proline and hydroxyl amino acids. Many threonine residues serve as glycosylation sites and the sugar residues protect the exposed LR from proteolysis **[55]**. The heavy glycosylation of the LR conveys rigidity while the alternating proline sequence conveys flexibility. The LR is flexible enough to prevent an NMR or X-ray structure of the full protein from being determined, but rigid enough to remain protruded from the CD after the CBM is removed **[56]**. Studies have shown that the length of the LR is important in determining hydrolytic activity on insoluble substrates **[51, 57]**.

1.5 Cellulase-Cellulose Reaction Characteristics

The complete hydrolysis of lignocellulosic materials by cellulolytic organisms involves the synergistic action of multiple sets of glycosyl hydrolases, where each set consists of functionally different enzymes that target the same carbohydrate. Accessibility and reactivity significantly influence the ability of these enzymes to

hydrolyze their substrates: More open (high specific surface area) and less crystalline (amorphous) substrates are easier to hydrolyze. Substrates containing lignin are difficult to hydrolyze because it can not be removed enzymatically and has also been shown to nonproductively bind cellulases. Many pretreatment methods are designed to remove lignin, increase surface area, and decrease crystallinity. Pretreatment methods can be biological, chemical, or physical and their application relies on an economic balance between the cost of increasing pretreatment "intensity" and the profits from the increased yield **[58]**.

Binding. The hydrolysis reaction is characterized by the rapid binding of cellulase enzymes to insoluble cellulose with a simultaneous reduction of enzyme concentration in the liquid phase. The association of the CBM's active-binding face with exposed cellulose at the liquid-solid interface is the primary cause for absorption. Inactivation of the attached CD does not affect the ability of the full-length protein to bind **[59],** and the binding of isolated CBMs is similar that of the full-length protein **[60]**. Binding occurs rapidly compared to the hydrolysis reactions, often within seconds or minutes. CBMs and full-length cellulases do not bind uniformly to the surface of cellulose and have been shown to prefer specific crystalline faces **[61, 62]**. CBMs are thought to have surface mobility, allowing the entire cellulase to diffuse along the surface **[63]**.

Based on the molecular mechanistic for cellulase binding, binding should be entirely reversibly because no covalent bonds are formed **[64]**. Yet experimental work has shown that in spite of multiple dilutions or washes with buffer solution, some cellulases do not desorb from their insoluble substrate **[65-67]**. Bothwell *et al*. showed compete binding reversibility with some cellulases but not with others **[68]**. Binding of isolated CBMs was shown to be completely reversible on crystalline cellulose, implicating the CD in the irreversibility of binding **[69]**. Although reversibility is still

debated, Langmuir binding isotherms, which assume reversible-equilibrium binding, are widely used to document cellulase binding on cellulose. They are useful in identifying and comparing maximum binding capacity for different combinations of enzyme and substrate.

Hydrolysis. Product formation from cellulose hydrolysis is often referred to as non-linear because it involves two characteristic phases: an initial rapid phase followed by a slow exhausted phase. Whereas enzyme binding usually occurs within seconds or minutes, hydrolysis occurs over hours and days, with the transition to the slow reaction phase occurring within the first few hours. Cellulase systems tend to hydrolyze cellulose faster and with higher conversion than their respective monocomponents. Endocellulases have been shown to cause surface pitting, while exocellulases have been shown to leave tracks along the surface **[70]**. Hydrolysis has been shown to change the hydrogen bond network **[71]** along with small changes in crystallinity. Longer term hydrolysis leads to fragmentation **[72, 73]**. In the final exhausted state, cellulases are often bound but with no product being formed.

There are many theories for the decrease in rate which can be grouped into two categories: those related to the enzyme and those related to the substrate **[10]**. Cellulase enzymes can deactivate/denature during hydrolysis, bind non-productively, or be inhibited by glucose and cellobiose. Studies that implicate enzymes as the cause for the decrease in rate will claim to remove "old" enzymes and stimulate additional hydrolysis by adding fresh enzyme **[66]**. At the same time, cellulose appears to have a fraction of the accessible substrate that is easily hydrolyzed, or easily hydrolysable fraction (EHF) **[74]**. Because more crystalline substrates are often more difficult to hydrolyze, many have assumed this EHF is exposed amorphous cellulose. Studies that implicate the substrate as the cause of rate reduction will claim that swapping "old"

enzyme for fresh does not increase conversion and that if additional substrate is added to an exhausted reaction, a spike in the reaction rate will occur.

1.6 Kinetics Models

A wealth of mechanistic models for cellulase/cellulose reaction kinetics are available in the literature **[75, 76]**. Models are termed "mechanistic" if they include the ability of the cellulase enzyme to bind to insoluble cellulose, as binding is considered to be a fundamental mechanism in the reaction process.

A prototypical minimal-mechanistic model for hydrolysis couples Langmuir enzyme binding with a single hydrolytic activity is as follows,

(1c) $$B = \frac{E_f}{E_f + K_d} \lambda C_s ;$$

(1b) $$\frac{1}{\alpha}\frac{dP}{dt} = k_{cat} B .$$

In the above mechanism, B (*mol/vol*) is the concentration of bound enzyme, E_f (mol/vol) is the concentration of free enzyme in the liquid phase, λ (*sites/mass*) is the concentration of binding-sites, C_S (*mass/vol*) is the concentration of insoluble cellulose in the reactor, K_d (*mol/vol*) is the dissociation constant for binding, P (*mol/vol*) is the hydrolysis products, α (*dimensionless*) is the amount of product per binding site hydrolyzed, and k_{cat} ($time^{-1}$) is the catalytic rate constant. An additional mass balance for substrate completes the equation set. The hydrolysis reaction is first order with respect to bound enzyme because binding is thought to be a perquisite for hydrolysis, and initial hydrolysis rates have by found to be directly proportion to the

bound enzyme concentration. The binding reaction is instantaneous, with the dissociation constant and binding site capacity inferred from low temperature binding isotherms, where the hydrolysis reaction is slowed dramatically.

The prototypical hydrolysis model described above is often "dressed up" with additional mechanisms. Production inhibition via rapid equilibrium is perhaps the most commonly added mechanism **[77-79]**. Both glucose and cellobiose have been shown to competitively inhibit cellulase enzyme by complexing in the enzyme's hydrolytic active site. Note that both enzyme binding and product inhibition can be modeled as non-equilibrium processes, in which case their governing algebraic equations become differential. Enzyme activity can also be reduced by introducing an inactivation mechanism, often a simple first order process **[79]**. More exotic enzyme related mechanism can be added, including multi-enzyme activities and synergism **[80]**. Because amorphous cellulose is easier to degrade than crystalline, dual-substrate models have been proposed where the substrate contains two different types of binding-sites each with separate binding affinities and hydrolytic activities. A fraction of the binding-sites may also be modeled as inert but able to bind the enzyme nonproductively. One must be careful when adding these mechanisms because they introduce additional degree-of-freedoms in the modeling process which may fortuitously improve the model's fit to data. While Zhang **[75]** proposed the name "semi-mechanistic" for these models, they are better known as "pseudo-homogenous" models because they assume the binding sites in the insoluble cellulose can be treated pseudo-molecular homogenous chemical species which interact with enzyme in solution via classical mass-action kinetics. This subtle simplification has powerful implications and it simplifies the kinetics problem to a homogenous one that can be described with ordinary differential equations (ODEs).

Pseudo-homogenous models essentially "black-box" the structure of the insoluble cellulose, which changes over the course of hydrolysis. Some researchers have considered adding mechanisms to pseudo-homogenous models which introduce new binding sites during hydrolysis or a rate constant that depends on time or conversion **[81]**. Although interesting, these mechanisms add an addition degree-of-freedom while revealing no new information about the structure of the cellulose. One way to introduce structural information into a kinetics model is to assume the cellulose particles have a fixed geometry in which a continuum of reactive binding sites exists. These distributed binding site models retain the simplifying concept of "binding sites" but distribute them within an insoluble phase, allowing one to consider reaction at individual points in the substrate. Therefore, they offer a distinct advantage over pseudo-homogenous models, in that they consider the true point-wise reaction rates rather than the homogenous mixing-cup rates. In applying such models one must include a penetration mechanism, wherein the cellulase enzymes can enter the substrate and degrade it. The inclusion of a penetration mechanism is complicated as one must consider time and position dependent changes in the porosity of the solid phase, the accessible surface area with the reacting phase, and diffusivity of the enzyme. Consequently, the limited availability of microstructural models for cellulose and how that structure changes over the course of hydrolysis have limited the availability of such models in the literature. Moreover, those that do exist are often ill-defined, which effectively reduces them to a pseudo-homogenous model.

Both pseudo-homogenous and distributed binding-site models are unable to describe the details of cellulose chain hydrolysis, particularly regarding endo vs. exo activity. In both cases, the molecular structure of the cellulose interface (including chain ends, crystalline faces, and amorphous surfaces) are lumped into generic locations, i.e. binding sites, where the enzyme simply attaches. A few models have

tried to avoid the concept of "binding sites" altogether by modeling the hydrolysis of individual cellulose chains. These models assume an ensemble of cellulose chains are attacked at interior bonds by endocellulases and attacked processively at their ends by exocellulases. The models are useful in describing the synergistic hydrolysis of soluble oligosaccharides or soluble cellulose derivatives such as CMC. To compensate for the fact that some chains are physically excluded from hydrolysis, an accessibility factor is often introduced, but this must be done in an *ad hoc* way to maintain a mathematically-tractable governing ODE set. Worse yet, these models have incorrectly assumed the binding of a single chain by a cellulase precludes its ability to bind another cellulase **[82]** and that cellulases cover a single cellulose bond **[83]**. Others have assumed bond cleavage specificity is random and independent of location of the cellulose chain relative to other cellulose chains **[84, 85]**. Indeed, tracking the location, accessibility, and hydrolysis histories of each cellulose chain is the most difficult hurdle to overcome without escalating the complexity of the governing mathematics. Kinetic Monte-Carlo (KMC) models have offered some advantages in dealing with these problems because they can simulate the location of individual chains in space and track their histories directly by simulating the action of various types of cellulases at a solid-liquid interface **[86-90]**. These models are elaborate and contain many degrees of freedom which the modeler must try to define from various experimental studies. Many theoretical obstacles must also be overcome in constructing such models. For instance, the exact location of every cellulose chain in 3-D space must be specified, along with a detailed mechanism for how the cellulase (modeled here as independent space-filling objects) associate with and cleave the exposed bonds.

1.7 Outline of Dissertation

This dissertation presents three enzyme kinetics studies, each with in-depth analysis of the governing equation sets. Through scaling we are able to identify criteria for quasi-steady-state (QSSA) approximations and provide approximate time-dependent solutions with perturbation techniques.

In our first study we revisit the well known Michaelis-Menten (MM) model for homogenous enzyme kinetics. After reviewing a widely accepted treatment by Lin and Segel **[91]**, we were motivated to generate our own uniformly-valid asymptotic solution using a combined regular and singular perturbation. Through a careful review of the literature, we found how each of many previous solutions matched our solution in various limiting regions of parameters space. Our scaling also revealed parameters which delineated where various QSSAs and Uniformly Valid Approximations (UVAs) relating to the MM problem were valid.

In our second study, we investigate the hydrolysis of p-nitrophenyl-cellobioside (pNP-G2), a modified cellulo-oligo substrate, by *Thermobifida fusca* Cel5Acd. Glycosides of nitrophenols are commonly used for studying cellulases because their hydrolysis yields an easily detectable nitrophenolic product **[92-94]**. The kinetics of pNP-G2 hydrolysis by Cel5A is complicated by the enzymes retaining mechanism, which allows for transglycosylation, obscuring in some cases the formation of p-nitrophenol (pNP) report. To account for this, we generated a mechanistic model for hydrolysis and transglycosylation in the enzyme's active site. The subsequent reaction network is identical to a previously described network and rate equation for similar enzyme-substrate systems. We improve upon this network by providing criteria that justified the use of the QSSA leading to the rate equation and further justified its integration to yield a solution for reporter product in time, which

we show fits well to our experimental data. Our analysis can be extended to similar systems with retaining glycosyl hydrolases acting on nitrophenyl-glycosides.

In our third study, we present a distributed binding-site model for the hydrolysis of a dilute suspension of slender, dense, cylindrical cellulose fibers by cellulase enzymes. We show, through a comparison of the characteristic timescales for diffusion and hydrolysis that an enzyme gradient in the reacting zone is unlikely, and that for our idealized substrate, the reacting zone in the fibers will grow at a constant rate. After scaling the governing PDE set, we are able to generate an approximate solution using a regular perturbation, which assumes a small relative reduction in the liquid phase enzyme concentrations occurs due to binding. The resulting solution reveals the time-dependent liquid phase concentration of bound enzyme and product cannot be used to identify the corresponding concentration profile in the reacting fiber (and the dominant underlying penetration mechanism.) More importantly, we show that application of pseudo-homogenous models to the bulk liquid phase concentration curves results in kinetics parameters that only relate directly to the true underlying intrinsic kinetics parameters in certain limiting circumstances. The study provides quantifiable evidence that the liquid-phase enzyme binding and product formation curves are not amendable to simple interpretation – either for identifying how cellulose microstructure changes or for elucidating intrinsic kinetics parameters. One must include a reliable penetration model and mechanism or provide additional experimental evidence that some limiting case scenario exists.

Finally, we present in an Appendix our collaborative work involving a novel electrospun cellulose (ESC) substrate. Here we made the first-ever attempt to assemble insoluble cellulose fibers with uniform, systematically-controlled microstructural properties.

CHAPTER 2

A NEW PERTURBATION SOLUTION TO THE MICHAELIS-MENTEN PROBLEM*

*Reproduced in part with permission from Dingee JW & Anton AB (2008) "A New Perturbation Solution to the Michaelis-Menten Problem," *AIChE Journal* **54** (5), 1344-1357 © 2008 American Institute of Chemical Engineers (AIChE).

2.1 Introduction

Homogeneous, enzyme-catalyzed reactions nearly always occur via a two-step process known as the "Michaelis-Menten" (MM) mechanism **[95-97]**. In the first elementary reaction, the enzyme E attaches reversibly to the substrate molecule S to form a complex C, i.e.

$$(1) \qquad E + S \underset{k_{-1}}{\overset{k_1}{\leftrightarrow}} C ;$$

and in the second elementary reaction, the bound enzyme converts the substrate irreversibly to product P and releases it, which returns the enzyme to solution, i.e.,

$$(2) \qquad C \overset{k_2}{\rightarrow} P + E .$$

The parameters k_1 (*mol/vol-time*), k_{-1} ($time^{-1}$), and k_2 ($time^{-1}$) are the kinetic rate constants for each elementary reaction as shown **[98]**. One frequently seeks to measure these rate constants (or combinations thereof) as accurately as possible for

various enzyme/substrate combinations to compare their performance or for use in simulations of complex, multi-step, multi-component biochemical processes **[99, 100]**.

Four ordinary differential equations are required to describe the time-dependent concentrations $E(t)$, $S(t)$, $C(t)$, and $P(t)$ in a reactor where an MM reaction is occurring:

(3a) $$\frac{dE}{dt} = -k_1\left(ES - K_M C\right);$$

(3b) $$\frac{dS}{dt} = -k_1\left(ES - K_D C\right);$$

(3c) $$\frac{dC}{dt} = k_1\left(ES - K_M C\right);$$

(3d) $$\frac{dP}{dt} = k_1 KC\ .$$

From this point forward we use combined kinetics parameters that are preferred by the biochemistry community: $K = \frac{k_2}{k_1}$ is the Van Slyke-Cullen constant **[101]**; $K_D = \frac{k_{-1}}{k_1}$ is the dissociation equilibrium constant of the enzyme-substrate complex; and $K_M = K + K_D = \frac{k_{-1} + k_2}{k_1}$ is the MM constant **[98]**. Each of these has units of concentration (*mol/vol*) and can range between zero and infinity for different enzyme-substrate combinations. They quantify in some sense the relative tendency of enzyme and substrate to distribute among initial, intermediate, and final states during the reaction. Typically, the reaction is carried out in a closed batch reactor charged with an initial concentration of enzyme (E_T) and substrate (S_T), in which case there are two algebraic mass constraints:

(4a) $$E_T = E + C\ ;$$

(4b) $S_T = S + P + C$.

Equations (3) and (4) can be combined into two independent ordinary differential equations, which, when solved with the initial conditions $E(t=0) = E_T$, $S(t=0) = S_T$, $C(t=0) = 0$, and $P(t=0) = 0$, give the time-dependent concentrations of all species in the reactor. The solution of these equations is a function of five parameters: the rate constants k_1, k_{-1}, and k_2 (or the combinations K, K_D, and K_M), which one is typically trying to measure in MM experiments, and the initial concentrations E_T and S_T, which one can vary in experiments to ensure good data are collected.

Unfortunately, the governing equations (3) are nonlinear and coupled, so no exact, closed-form solution is available, which makes it difficult to extract estimates of the kinetic parameters from curve-fits to concentration-vs.-time data. Various simplifications and approximations have been proposed to circumvent this difficulty, each of which is only accurate for some limited range of conditions. We will discuss and compare them in some detail later in this chapter after we present and justify a new, more accurate perturbation solution to the MM equations.

2.2 Scaling and Solution

We begin by combining (3) and (4) to recast the problem in terms of $C(t)$ and $P(t)$, which we find to be the most expedient pair of concentrations to use for the ensuing analysis:

(5a) $\dfrac{dC}{dt} = k_1\left[(E_T - C)(S_T - P - C) - K_M C\right]$ with $C(0) = 0$;

(5b) $\frac{dP}{dt} = k_1 KC$ with $P(0) = 0$.

Since the intermediate complex must form according to reaction (1) before the substrate can be converted to product according to reaction (2), there are two natural timescales in this coupled, nonlinear pair of equations. If the two timescales are widely disparate, i.e., complex forms quickly and product forms slowly, which we will show to be unavoidably the case here, and if the nonlinearities are relatively weak, the MM equations can be solved to arbitrary accuracy with a rigorous approximation method known as "perturbation theory" **[91, 102, 103]**. Perturbation methods are ubiquitous in chemical engineering, including applications in fluid mechanics **[104, 105]**, catalysis **[106]**, combustion **[107, 108]**, semiconductors **[109]**, and control theory **[110, 111]**.

To apply perturbation theory, the first and most crucial move is to "scale" **[91]** the dependent variables, in this case C and P, with appropriate combinations of the constant parameters in the equations to generate new, dimensionless variables that are certain to be $O(1)$ ($O(x)$="order of x" in an asymptotic sense. *Cf.* **[103]**, chapter 1.3.) over the full range of the independent variable, in this case $0 \le t < \infty$. When the scaling causes a small dimensionless parameter, which we will call δ, to appear as a coefficient of nonlinear terms in the equation that are otherwise $O(1)$ functions of the dependent variables, a *regular* perturbation is indicated. In this situation the equations can be solved by orders in δ with a perturbation series. When the scaling procedure reveals two different timescales that can be used to scale the independent variable t, a *singular* perturbation is indicated. The ratio of the short timescale to the long timescale appears as a small dimensionless parameter, which we will call ε. One casts the equations in terms of the short-time variable and solves them by orders in ε with a

perturbation series, and then recasts them in terms of the long-time variable and solves them again by orders in ε with a perturbation series. The "inner" solution for short time and the "outer" solution for long time are then matched asymptotically in the intermediate region where the two timescales overlap. Our analysis of the MM problem is unusual in that perturbations of both types appear together and must be solved simultaneously. For a more detailed discussion of these methods and their applicability to problems of this type, we refer the reader to any of many fine textbooks on the subject **[91, 102, 103]**.

Perturbation methods offer several advantages over numerical techniques for problems of this type. First and foremost, the leading-order, $\delta, \varepsilon \to 0$ portion of the solution, which is often easy to find, describes the exact solution to sufficient accuracy for many uses and reveals its dominant behavior. Furthermore, a properly scaled perturbation solution is uniformly valid over the entire range of the independent variable and not subject to the "stacking up" of numerical integration errors. Because of the ordering scheme, the equations for higher-order corrections are certain to be linear and soluble, and because the scaling ensures the dimensionless variables are $O(1)$, the difference between the perturbation and exact solutions at every point is certain to be at most $O(\delta^n$ or $\varepsilon^n)$, where n is the order of the first neglected term in the perturbation expansion. One retrieves an orderly prescription for finding an analytical solution with arbitrary and quantifiable accuracy.

We begin our analysis by transforming the equations (5) to scaled, dimensionless forms and identifying the characteristic short and long timescales. Note first that equation (5a) has a constant "source" (+) term $k_1 E_T S_T$ on the right hand side. The correct choice of scales must drive this term to one for $O(1)$ balancing against the remaining terms. Note also that the form of (5) ensures that $C(t)$ rises from zero at $t = 0$, passes through a maximum, and then decreases to zero as $t \to \infty$, whereas $P(t)$

increases monotonically from zero at $t = 0$ to S_T as $t \to \infty$. Let the constant parameter β be a leading-order estimate of the maximum in $C(t)$, and define dimensionless, $O(1)$ dependent variables $x = \frac{C}{\beta}$ and $y = \frac{P}{S_T}$. Substituting x and y as described above and dividing by $k_1 E_T S_T$ to convert the source term to one converts (5a) to the following dimensionless form:

$$(6) \qquad \left(\frac{\beta}{k_1 E_T S_T}\right)\frac{dx}{dt} = 1 - \left(\frac{E_T + S_T + K_M}{E_T S_T}\right)\beta x + \left(\frac{\beta^2}{E_T S_T}\right)x^2 - \left[1 - \left(\frac{\beta}{E_T}\right)x\right]y .$$

Since we anticipate that complex forms rapidly relative to product, we solve for β by setting $y \cong 0$ in this equation and considering the $t \to \infty$ limit $\left(i.e., \frac{dx}{d\tau} \to 0\right)$ of the remaining terms. Pair-wise balancing of the remaining terms on the right-hand side confirms that the constant and the linear term are dominant, and the term quadratic in x is subdominant (*cf.* **[91]**, chapter 9.1; *cf.* **[102]**, chapter 3.4), from which we find

$$(7) \qquad \beta = \frac{E_T S_T}{E_T + S_T + K_M} .$$

Note that this choice of β sets the coefficient of the term linear in x equal to one, ensuring an $O(1)$ balance of the dominant source and sink terms.

Since the scaled variable x accounts for complex formation, the dominant process on the short timescale, and since equation (6) is dimensionless as shown, one can immediately identify the coefficient of the derivative as the scaling factor for t on the short timescale. Consequently, we define a dimensionless independent variable $\tau = \frac{t}{\theta_C}$, where

$$(8) \qquad \theta_C = \frac{\beta}{k_1 E_T S_T} = \frac{1}{k_1\left(E_T + S_T + K_M\right)}$$

is the characteristic (short) time for complex formation. Note the rational dependence on kinetics parameters evident in θ_C: it is the ratio of the amount of complex to be produced on the short timescale (β) to a characteristic rate of complex formation ($k_1 E_T S_T$).

Now one is left with two groups of constants on the right-hand side of equation (6). For convenience we define two additional dimensionless parameters:

$$\gamma = \frac{\beta}{E_T} = \frac{S_T}{E_T + S_T + K_M}; \tag{9}$$

$$\delta = \frac{\beta^2}{E_T S_T} = \frac{E_T S_T}{\left(E_T + S_T + K_M\right)^2}. \tag{10}$$

These dimensionless parameters have useful physical meanings. Since β is a leading-order estimate of the maximum complex concentration, γ is clearly a leading-order estimate of the fraction of enzyme initially charged to the reactor that forms complex on the timescale θ_C. Note that $0 < \gamma < 1$ for all possible choices of the parameters E_T, S_T, and K_M $\left(0 \leftrightarrow \infty\right)$. The meaning of the parameter δ is less obvious but can be gleaned from the following consideration. If the MM reaction is conducted in a flow reactor with continuous replenishment of enzyme at a fixed concentration E_T, so that the mass constraint for total enzyme, equation (4a), does not enter the problem, the term δx^2 does not appear in equation (6), or alternatively, $\delta = 0$. Thus δ quantifies the extent to which equilibrium of the complex-formation reaction (1) is perturbed by the fact that the reaction is conducted in a closed vessel with a finite supply of enzyme. Note by inspection of (10) that $0 < \delta < \frac{1}{4}$ for any possible combination of E_T, S_T, and K_M, so the nonlinear term $\delta x^2 << 1$ always, which will allow us to treat it as a regular perturbation.

When one propagates the same scaling to equation 5(b), one finds

$$(11) \qquad \frac{dy}{d\tau} = \left(\frac{\theta_C k_1 K \beta}{S_T} \right) x .$$

This equation must reveal the long timescale over which product formation is the dominant process. Recall we set $y \cong 0$ when analyzing equation (6) to identify the dominant behavior on the short timescale. This condition is met if the dimensionless coefficient on the right hand side of equation (11) is small and represents the ratio of the short and long timescales. Consequently, we identify

$$(12) \qquad \varepsilon = \frac{\theta_C k_1 K \beta}{S_T} = \frac{\theta_C}{\theta_P} = \frac{K E_T}{\left(E_T + S_T + K_M \right)^2}$$

as the singular perturbation parameter for this problem. The form of (12) ensures that $0 < \varepsilon < \frac{1}{4}$ for any possible combination of E_T, S_T, K, and K_M $\left(= K + K_D \right)$, and combining (10) and (12) reveals that $\delta + \varepsilon < \frac{1}{4}$, which is indeed a fortunate circumstance, as we will show in the Appendix. It follows immediately that

$$(13) \qquad \theta_P = \frac{S_T}{k_1 K \beta} = \frac{E_T + S_T + K_M}{k_1 K E_T}$$

is the characteristic (long) timescale for product formation. Note that θ_P is the ratio of the amount of product to be formed on the long timescale (S_T) to a characteristic rate of product formation ($k_1 K \beta$). Finally, one can identify a "stretched" form of the independent variable $T = \frac{t}{\theta_P} = \varepsilon \tau$ that is appropriate for the long timescale, where product formation is the dominant process.

Incorporating everything to this point, the short-time $\left(\tau = \frac{T}{\varepsilon} \right)$ or "inner" representation of equations (5a-b) becomes

(14a) $\dfrac{dx}{d\tau} = 1 - x + \delta x^2 - (1 - \gamma x) y$ with $x(0) = 0$;

(14b) $\dfrac{dy}{d\tau} = \varepsilon x$ with $y(0) = 0$.

Note that x is constrained to be $O(1)$ and y is constrained to be $O(\varepsilon)$ on the short timescale where τ is $O(1)$, since $\delta < \dfrac{1}{4}$ and $\gamma < 1$ always. The long-time $(T = \varepsilon\tau)$ or "outer" representation of the problem, distinguished here by use of capital letters for the dependent variables, becomes

(15a) $\varepsilon\left(\dfrac{dX}{dT}\right) = 1 - X + \delta X^2 - (1 - \gamma X) Y$;

(15b) $\dfrac{dY}{dT} = X$.

Now one observes for the same reasons as before that X and Y are both constrained to be $O(1)$ on the long timescale, where T is $O(1)$. The outer equations have no initial conditions; rather, they must asymptotically match the inner solutions in the region where the inner and outer solutions overlap.

The form of equation (15a) reveals a special physical significance of the singular perturbation parameter ε. Since X, Y, and T are $O(1)$ on the outer timescale, $\dfrac{dX}{dT}$ is $O(1)$ also, and this derivative vanishes from equation (15a) when $\varepsilon \to 0$, which makes (15a) an algebraic rather than a differential equation. This is the familiar "quasi-steady-state approximation" (QSSA) of chemical kinetics **[112-114]**. One concludes that the magnitude of ε quantifies the extent to which the QSSA is valid for

the complex concentration in the MM problem, since $X(T)$ is the dimensionless representation of $C(t)$. Recall we found earlier that $0 < \varepsilon < \frac{1}{4}$, which confirms that the QSSA $\frac{dC}{dt} \cong 0$ is *always* at least *approximately* valid for the MM problem **[115]**. Stated another way, the scaling and ordering scheme reveals that no matter what values of E_T and S_T are chosen for experiments and what values of the elementary rate constants k_1, k_{-1}, and k_2 obtain for the enzyme/substrate system under investigation, the relative errors in quantities extracted from MM data analyses that use the QSSA will be less than approximately 25%. A more thorough discussion of the QSSA for MM kinetics will be presented later in this paper.

Our scaling has successfully constrained the dependent variables x and y to be $O(1)$ in both the inner and outer representations, and this ensures that their perturbation representations will be orderly convergent for all possible values of the parameters δ, γ, and ε. The equations are now properly transformed for solution with a combined regular (parameter δ) and singular (parameter ε) perturbation series. We refer the reader now to the Appendix, where the solution is described in detail.

2.3 Limiting Forms of the Perturbation Solution

Once x and y have been determined by solving equations (14) and (15), the algebraic mass constraints (4) can be used to find $E(t)$ or $S(t)$. It is useful to recast this part of the problem in dimensionless form, too. The natural scales for $E(t)$ and $S(t)$ are E_T and S_T, of course, so we define additional dimensionless variables $w = \frac{E}{E_T}$ and $z = \frac{S}{S_T}$. In terms of these, the mass constraints (4) can be rearranged to read

(16a) $w = 1 - \gamma x$;

(16b) $z = 1 - \mu x - y$.

A new dimensionless parameter appears,

$$(17) \qquad \mu = \frac{\beta}{S_T} = \frac{E_T}{E_T + S_T + K_M},$$

which quantifies the fraction of substrate initially present that is converted into complex on the short timescale θ_C. Two ancillary dimensionless ratios can also be identified,

$$(18) \qquad \eta = \frac{K\beta}{E_T S_T} = \frac{K}{E_T + S_T + K_M};$$

$$(19) \qquad \phi = \frac{K_D \beta}{E_T S_T} = \frac{K_D}{E_T + S_T + K_M}.$$

These can be thought of as kinetic branching ratios. When $\eta \to 0$, complex is converted slowly to product, and when $\phi \to 0$, complex is returned slowly to enzyme and substrate. When both are small, $K_M << E_T + S_T$, thus complex accumulates to its maximum possible concentration [*cf.* equation (7) for β], and when both approach one, $K_M >> E_T + S_T$, thus complex does not accumulate relative to E_T or S_T. Clearly, $\gamma + \eta + \mu + \phi = 1$, but more importantly, the magnitudes of these serve to distinguish among several approximate solutions for the MM problem that have appeared previously in the literature. We will return to this in the next section 2.4.

Since all six dimensionless parameters we have identified – the ratios γ, μ, η, and ϕ, and the perturbation parameters δ and ε – come from only four physical quantities in the original MM problem – E_T, S_T, K, and K_D – there are additional relationships among them, for example,

(20) $\delta = \gamma\mu$, $\delta \le \gamma(1-\gamma)$, $\delta \le \mu(1-\mu)$, and $\delta \le \frac{1}{4}(1-\eta)^2$;

(21) $\varepsilon = \eta\mu$, $\varepsilon \le \frac{1}{4}(1-\gamma)^2$, $\varepsilon \le \mu(1-\mu)$, and $\varepsilon \le \eta(1-\eta)$.

We present these because they show that some limiting values ($\to 0$ or $\to 1$) of γ, μ, η, or ϕ translate directly into limiting forms of the perturbation solution.

There are 14 possible inequality relationships among the MM parameters E_T, S_T, K, and K_D, which we list in Table 2.1 with corresponding consequences for the dimensionless ratios and perturbation parameters. The limiting cases for γ, δ, and ε are clearly most important, because they appear in the original scaled equations (14) and (15), hence control the fundamental nature of the solution. The implications of limiting values for the other parameters η, μ, and ϕ are more subtle, but we will demonstrate briefly here and in more detail in the next section how they serve to differentiate various physical scenarios for the MM mechanism.

A key concern in previous analyses of the MM mechanism has been determining when can one assume the complex concentration remains relatively low, so that either the free enzyme concentration remains nearly constant, or the substrate initially charged to the system exists thereafter only as product or as substrate remaining to be converted. Each of these limiting circumstances simplifies the mass balances (3). Under what circumstances can one expect them to occur?

Consider first the case where the free enzyme concentration remains effectively constant, or $E(t) = E_T - C(t) \cong E_T$. One intuitively expects this to occur when enzyme is present in large excess, i.e. $E_T >> S_T$, however the dimensionless form of the enzyme mass constraint, equation (16a), shows that this is actually achieved, i.e. $w \cong 1$, when $\gamma = \frac{\beta}{E_T} \to 0$. Rows (a) through (g) of Table 2.1 list six

Table 2.1 Limiting Values of the Dimensionless MM Parameters. The limiting values are consequences of inequality relationships among the four independent MM parameters E_T, S_T, K and K_D. (Recall: $K_M = K + K_D$.) The dimensionless ratios γ, η, μ, and ϕ are defined in terms of the MM parameters by equations (9), (18), (17), and (19), and the perturbation parameters δ and ε are defined by equations (10) and (12), respectively. The columns labeled "QSSA" and "UVA" list references to quasi-steady-state and uniformly-valid approximations, respectively, that are accurate for each case. Tzafriri and Edleman's uniformly-valid solution **[116]** is accurate in all situations to within errors of $O(\varepsilon)$, and the perturbation solution we present here is accurate in all situations to within errors of $O(\delta + \varepsilon)^2$.

Case	Description	γ	η	μ	ϕ	δ	ε	QSSA	UVA
(a)	$E_T >> S_T + K_M$	$\to 0$	$\to 0$	$\to 1$	$\to 0$	$\to 0$	$\to 0$	**[117-119]**	**[120]**
(b)	$E_T + K_M >> S_T$	$\to 0$	<1	<1	<1	$\to 0$	$< 1/4$		**[120]**
(c)	$E_T + K_D >> S_T + K$	$\to 0$	$\to 0$	<1	<1	$\to 0$	$\to 0$	**[117-119]**	**[120]**
(d)	$E_T + K >> S_T + K_D$	$\to 0$	<1	<1	$\to 0$	$\to 0$	$< 1/4$		**[120]**
(e)	$K >> E_T + S_T + K_D$	$\to 0$	$\to 1$	$\to 0$	$\to 0$	$\to 0$	$\to 0$	**[118, 121, 122]**	**[120]**
(f)	$K_D >> E_T + S_T + K$	$\to 0$	$\to 0$	$\to 0$	$\to 1$	$\to 0$	$\to 0$	**[118, 119, 121, 122]**	**[120]**
(g)	$K_M >> E_T + S_T$	$\to 0$	<1	$\to 0$	<1	$\to 0$	$\to 0$	**[118, 121, 122]**	**[120]**
(h)	$S_T + K_M >> E_T$	<1	<1	$\to 0$	<1	$\to 0$	$\to 0$	**[118, 121, 122]**	**[117, 123, 124]**
(i)	$S_T >> E_T + K_M$	$\to 1$	$\to 0$	$\to 0$	$\to 0$	$\to 0$	$\to 0$	**[118, 119, 121, 122]**	**[117, 123, 124]**
(j)	$S_T + K_D >> E_T + K$	<1	$\to 0$	$\to 0$	<1	$\to 0$	$\to 0$	**[118, 119, 121, 122]**	**[117, 123, 124]**
(k)	$S_T + K >> E_T + K_D$	<1	<1	$\to 0$	$\to 0$	$\to 0$	$\to 0$	**[121, 122]**	**[117, 123, 124]**
(l)	$E_T + S_T + K_D >> K$	<1	$\to 0$	<1	<1	$< 1/4$	$\to 0$	**[118, 119]**	
(m)	$E_T + S_T >> K_M$	<1	$\to 0$	<1	$\to 0$	$< 1/4$	$\to 0$	**[118, 119]**	
(n)	$E_T + S_T + K >> K_D$	<1	<1	<1	$\to 0$	$< 1/4$	$< 1/4$		

different limiting relationships among the MM parameters that cause $\gamma \to 0$. Only row (a) accounts for the intuitively obvious excess-enzyme scenario, because $\gamma \to \frac{S_T}{E_T} \to 0$ when $E_T >> S_T + K_M$, but rows (b-g) describe five other scenarios where $\gamma \to 0$ can be realized without firm constraints on the relative magnitude of E_T and S_T.

Consider instead the case where the initial substrate is present only as remaining substrate or as product, or $S(t) = S_T - C(t) - P(t) \cong S_T - P(t)$. One intuitively expects this to occur when substrate is present in large excess, i.e. $S_T >> E_T$, however the dimensionless form of the substrate mass constraint, equation (16b), shows that this is achieved, i.e. $z \cong 1 - y$, when $\mu = \frac{\beta}{S_T} \to 0$. Only row (i) accounts for the intuitively obvious excess-substrate scenario, because $\mu \to \frac{E_T}{S_T} \to 0$ when $S_T >> E_T + K_M$, but rows (e-h) and (j-k) describe five other scenarios where $\mu \to 0$ can be realized without firm constraints on the relative magnitude of E_T and S_T.

2.4 The Original Michaelis-Menten Equation and Related Quasi-Steady-State Approximations

The QSSA is a clever "trick" for simplifying the complicated sets of differential equations that are ubiquitous in chemical kinetics analyses **[125, 126]**. The premise of the QSSA is that some intermediate chemical species are consumed so rapidly after they form in a reaction mechanism, they can not accumulate to an appreciable extent; consequently, time-derivatives of their concentrations are nearly zero. This simplification converts differential equations into algebraic equations and reduces the order and complexity of the mathematical problem at hand, but at the expense of completeness, because removing time-derivatives from the problem

ensures that the solution cannot meet the initial conditions for all species. One typically uses chemical intuition and insight to decide when the QSSA might be valid, but the best decider is a rigorous scaling analysis like we have presented here, as it reveals what relationships among parameters and initial conditions actually justify a QSSA.

The "MM approximation," which is included in the discussion of enzyme kinetics in any biochemistry textbook **[95, 96]**, is often cited as a prototypical example of the utility of the QSSA **[113]**. One begins by requiring that all MM experiments are conducted with excess substrate, i.e. $\frac{E_T}{S_T} \to 0$, so $C(t) << S(t) + P(t)$ and $S(t) + P(t) \cong S_T$ are assured, and one may safely assume $\frac{dC}{dt} \cong 0$. This particular set of conditions and assumptions is also referred to in the literature as the "standard QSSA" or sQSSA of MM kinetics **[121]**. Substituting (4a) into (3c) and simplifying via the QSSA gives

$$C \cong \frac{E_T S}{S + K_M}. \tag{22}$$

Combining this with (3d) and (4b) generates the so-called "MM equation":

$$\frac{dP}{dt} = -\frac{dS}{dt} \cong \frac{V_{\max} S}{S + K_M}; \tag{23}$$

where $V_{\max} = k_1 K E_T$ is the maximum attainable rate for a given enzyme concentration.

The sQSSA can be better understood if reexamined in the context of our scaled representation of equation set (3). The assumption that an initial complex-forming transient has passed implies that a time $t >> \theta_C$ passes before data are collected, so that the outer equations (15) apply. The assumption $\frac{dC}{dt} \cong 0$ is equivalent to $\varepsilon \to 0$ in (15a), and the experimental constraint $\frac{E_T}{S_T} \to 0$ is equivalent to $\mu \to 0$ in (17),

which ensures $C(t) << S(t) + P(t)$ according to (16b). But, because $\varepsilon = \eta\mu$ according to (21), the single criterion $\mu \to 0$ is sufficient to ensure that the sQSSA is valid. The sQSSA criteria $\frac{E_T}{S_T} \to 0$ first proposed by Laidler **[122]**, $\frac{E_T}{K_M} \to 0$ proposed more recently by Seshadri and Fritch **[123]**, and $\frac{E_T}{S_T + K_M} \to 0$ advocated by Segel & Slemrod **[117]** are all special cases of the more general criterion $\mu \to 0$. All of the limiting scenarios that cause $\mu \to 0$ and thereby substantiate the sQSSA are enumerated in rows (e) through (k) of Table 2.1.

The first integral solution to the MM kinetic problem was derived by the solving the MM equation (23) with the initial condition $S(t=0) = S_T$, which gives **[97]**

$$(24) \qquad \frac{S}{S_T} = 1 - \left(\frac{V_{\max} t}{S_T}\right) - \left(\frac{K_M}{S_T}\right) \ln\left(\frac{S}{S_T}\right).$$

The criterion $\mu \to 0$ that justifies this approach also ensures that the $O(\delta)$ and $O(\varepsilon)$ terms in our perturbation solution become vanishingly small (*cf.* Table 2.1); consequently, the integrated MM equation is identical to our leading-order solution $Y_0(T)$, equation (A14a), when $\mu \to 0$.

Experimentalists often collect $P(t)$ vs. t data of for a relatively short time-period and extract values of MM parameters by comparing the their data to the initial-rate form of (23), which is

$$(25) \qquad V_0 = \left.\frac{dP}{dt}\right|_{t=0} = -\left.\frac{dS}{dt}\right|_{t=0} = \frac{V_{\max} S_T}{S_T + K_M}.$$

The additional assumption $S \cong S_T$ in this treatment is equivalent to $Y(T) \to 0$ in our dimensionless representation, which can occur only if insufficient time has passed to penetrate significantly into the outer timescale, or alternatively $t << \theta_P$. Thus one

requires both $t >> \theta_C$ and $t << \theta_P$ to use equation (24), which can only be accomplished when the short and long timescales are well separated, or $\varepsilon \to 0$. Once again we see that $\mu \to 0$ is what substantiates even this more restrictive assumption.

Several more sophisticated approximate solutions of equation set (3) have appeared in the literature, including singular perturbations similar to ours but with $\frac{E_T}{S_T}$ **[91, 124, 127, 128]**, $\frac{E_T}{K_M}$ **[123]**, and $\frac{E_T}{S_T + K_M}$ **[117]** as the perturbation parameters. These are not QSSAs, but rather are "uniformly-valid" approximations (UVAs) that meet the initial conditions for all species. Note that $\mu \to 0$ when any of these ratios is small, thus $\delta, \varepsilon \to 0$ in our representation; consequently, they duplicate our leading-order solution and the integrated MM equation (24) for $S(T)$ when $\mu \to 0$. However, they differ significantly from our solution and from the exact solution when μ is not small. We will return to this in the next section, where we compare phase plots for various approximate solutions of the MM problem.

Recall that the sQSSA uses $\frac{E_T}{S_T} \to 0$ to ensure $C(t)$ is small, so one may safely assume $\frac{dC}{dt} \cong 0$. Borghans *et al.* realized the QSSA arises fundamentally from a mismatch of timescales and should apply regardless of the magnitude of $C(t)$ **[118]**. Substituting (4b) into (3c) to eliminate $S(t)$ and setting $\frac{dC}{dt} \cong 0$ gives $P(t)$ as a quadratic function of $C(t)$, which can in principle be inverted and used to integrate (3d) for $P(t)$. They effectively linearized the C vs. P relation with a Pade' approximation, and they used their simplified equations to identify correctly the same short and long timescales we identify by scaling equation (5) directly. They recognized that the ratio of these timescales, which is exactly the same as our equation (12) for ε, is less than unity for all possible circumstances, but they did not exploit this to develop a uniformly valid solution for the full problem. They named their approach the "total QSSA" (tQSSA), which is an improvement over the sQSSA because it adds at least some effect of the terms in the mass balances (3) that are quadratic in $C(t)$.

Because these are the same terms that appear as δx^2 in our scaled equations (14a) and (15a), respectively, and because the tQSSA is valid for $\varepsilon \to 0$, we conclude that the tQSSA extends the sQSSA to include entries (a), (c), (l) and (m) in Table 2.1.

A recent uniformly valid solution from Tzafriri and Edelman **[116]** is worthy of particular attention here, as it is similar to ours in some ways, yet has noteworthy differences. To facilitate a lucid comparison, we will recast their analysis in terms of our own dimensionless parameters.

Tzafriri **[115]** used intuition and physical arguments to correctly identify the same disparate timescales – θ_C (8) and θ_P (13) – which we identified with a routine scaling analysis, but he included the small quadratic term δx^2 in the leading-order scaling, whereas we used a dominant-balance argument to exclude it. Consequently, he obtained the following scaling constants and singular perturbation parameter (denoted by *), which, when expanded for small δ, are identical to our equations (7), (8), (13), and (12) in leading order:

$$\text{(26a)} \quad \beta^* = \beta \frac{1-\sqrt{1-4\delta}}{2\delta} \approx \beta\left[1+\delta+O\left(\delta^2\right)\right];$$

$$\text{(26b)} \quad \theta_C^* = \frac{\beta^*}{k_1 E_T S_T} = \frac{\theta_C}{\sqrt{(1-4\delta)}} \approx \theta_C\left[1+2\delta+O\left(\delta^2\right)\right];$$

$$\text{(26c)} \quad \theta_P^* = \frac{S_T}{k_1 K \beta^*} = \frac{2\delta}{1-\sqrt{1-4\delta}}\theta_P \approx \theta_P\left[1-\delta+O\left(\delta^2\right)\right];$$

$$\text{(26d)} \quad \varepsilon^* = \frac{\theta_C^*}{\theta_P^*} = \frac{1-\sqrt{(1-4\delta)}}{2\delta\sqrt{(1-4\delta)}}\varepsilon \approx \varepsilon\left[1+3\delta+O\left(\delta^2\right)\right].$$

Tzafriri and Edelman **[116]** solved the $O(1)$ inner equations, their equivalent of (14) with $\varepsilon = 0$, without resorting to a regular perturbation in δ, hence they obtained the Ricatti solution we give as equation (A4). They solved the nonlinear $O(1)$ outer equations, their equivalent of (15) with $\varepsilon = 0$, with an ingenious variable transformation and then constructed a uniformly valid solution by multiplying the inner and outer solutions together rather than matching them asymptotically. Their *ad hoc* solution accounts accurately for leading and higher-order effects of the quadratic nonlinearity quantified by δ, but only the $O(1)$ effect of the two timescales quantified by ε^* in (26d), hence its error, as quantified by the first neglected term, is $O(\varepsilon^*) \cong O(\varepsilon)$. Nonetheless, Tzafriri and Edelman's treatment is a dramatic improvement over all that preceded it, because it is the first uniformly valid solution to the full MM problem that is based on correct choices of the inner and outer timescales, which ensures it is approximately valid for any possible combination of the MM parameters.

Finally, a separate QSSA treatment has been proposed for equation (3b) that is referred to in the literature as the "reverse QSSA" (rQSSA) **[117]**. Setting $\frac{dS}{dt} \cong 0$ in (3b) yields

$$C = \frac{E_T S}{S + K_D}. \tag{27}$$

The implications of this approximation become clearer if one inspects the corresponding outer equations in dimensionless form. Scaling equations (3b) and (3c) appropriately and using (16a) to eliminate $W = \frac{E}{E_T}$ yields the following:

$$\eta\left(\frac{dZ}{dT}\right) = -\left[(1 - \gamma X)Z - \phi X\right]; \tag{28a}$$

(28b) $$\varepsilon\left(\frac{dX}{dT}\right) = (1-\gamma X)Z - (\phi+\eta)X \,.$$

Since $Z(T)$ is the dimensionless representation of $S(t)$, equation (28a) reveals that the criterion for validity of the rQSSA is $\eta \to 0$, but equation (21) and Table 2.1 verify that this also ensures $\varepsilon \to 0$ in (28b). We conclude in agreement with others that the rQSSA is more restrictive than the tQSSA, or stated another way, the assumption $\frac{dS}{dt} \cong 0$ must have associated with it the assumption $\frac{dC}{dt} \cong 0$. Segel and Slemrod proposed the criterion $E_T >> K_M$ to justify the rQSSA **[117]**, whereas Schnell and Maini argued that assuming $\frac{dC}{dt} \cong 0$ was not necessary, but $E_T >> K$ and $E_T >> S_T$ were **[119]**. The former corresponds to $\eta, \phi, \varepsilon \to 0$ in our dimensionless representation [Table 2.1, row (m)], and the latter corresponds to $\gamma, \eta, \delta, \varepsilon \to 0$ [Table 2.1, rows (a), (c), and (f)]. Rows (i), (j), and (l) in Table 2.1 identify three other, previously unidentified relationships among the MM parameters that also cause $\eta \to 0$ and therefore render the rQSSA valid.

2.5 Quantitative Comparison of Solutions

First we present plots of $X(T)$ and $Y(T)$ that show the constituent parts of the perturbation solution and compare the assembled solutions to numerical solutions of the scaled outer equations (15).

Recall that the scaled perturbation solutions are assembled from three functions, *e.g.*, $X(T) = X_0(T) + \delta X_\delta(T) + \varepsilon X_\varepsilon(T)$. The top panel of Figure 2.1 compares $X_0(T)$, $X_\delta(T)$, and $X_\varepsilon(T)$ from equations (A14a) and (A14b), and the bottom panel compares $Y_0(T)$, $Y_\delta(T)$, and $Y_\varepsilon(T)$ from equations (A14a) and (A14c) for representative values of the parameters $\gamma = 0.333$ and $\varepsilon = 0.110$. These functions are

independent of δ, but depend parametrically on γ because it enters the governing equations (A6) and (A8) at every order, and they also depend parametrically on ε because each perturbation function includes asymptotically matched contributions from both the short $\left(\tau = \frac{T}{\varepsilon}\right)$ and long $\left(T = \varepsilon\tau\right)$ timescales (*cf.* section 2.8). We present them in terms of the outer time-variable T because it is the $O(1)$ timescale for product formation.

The general shapes of the $O(\varepsilon)$ functions in Figure 2.1 are easy to interpret in the context of the MM reaction mechanism. Recall that ε is the ratio of the inner to outer timescales according to equation (12). As ε grows from zero, the short-time process, i.e. formation of complex, accelerates relative to the long-time process, i.e. formation of product. Consequently, one expects $\varepsilon > 0$ to be evidenced as an increase in the amplitude of the complex curve $X(T)$ and a decrease in the amplitude of the product curve $Y(T)$ for short times, to be made up by opposite trends at long times. Notice how $X_\varepsilon(T)$ first goes positive and then negative, and $Y_\varepsilon(T)$ does just the opposite, so that the $O(\varepsilon)$ corrections $\varepsilon X_\varepsilon(T)$ and $\varepsilon Y_\varepsilon(T)$ contribute to $X(T)$ and $Y(T)$ in exactly the manner just described. Unfortunately, the general shapes of the $O(\delta)$ functions in Figure 2.1 are not amenable to such simple interpretation. One must analyze carefully the equations (A6c,d) and (A8c,d) that generated them.

Figure 2.2 compares the assembled perturbation solutions to "exact" numerical solutions **[129, 130]** of the outer equations (15) for $\gamma = 0.333$, $\delta = 0.111$, and $\varepsilon = 0.110$. The top panel shows $X(T)$ (scaled complex) and the bottom panel shows $Y(T)$ (scaled product). Recall that the definitions of δ and ε, equations (19) and (12), ensure that $\delta + \varepsilon < \frac{1}{4}$ for any combination of the MM parameters; consequently, the choice of $\delta + \varepsilon = 0.221$ for this calculation represents a near-worst-case scenario for our perturbation solution. Nonetheless, the agreement between the perturbation and exact solution is excellent, but this should be expected, since the scaling and ordering

Figure 2.1: Constituent Parts of the Perturbation Solution. The curves in the top panel compare $X_0(T)$, $X_\delta(T)$, and $X_\varepsilon(T)$ from equations (A14a) and (A14b), and the curves in the bottom panel compare $Y_0(T)$, $Y_\delta(T)$, and $Y_\varepsilon(T)$ from equations (A14a) and (A14c), all for $\gamma = 0.333$ and $\varepsilon = 0.110$.

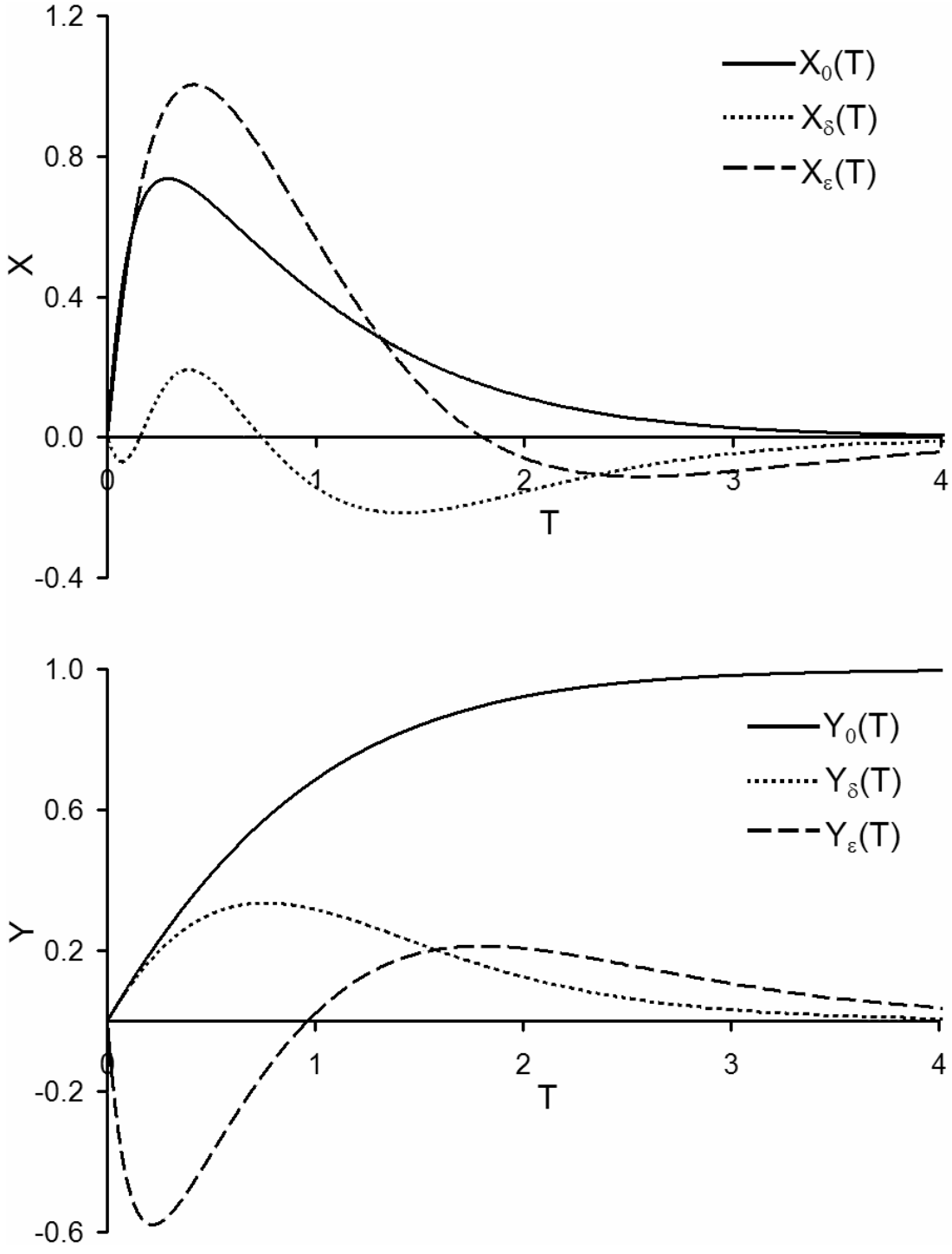
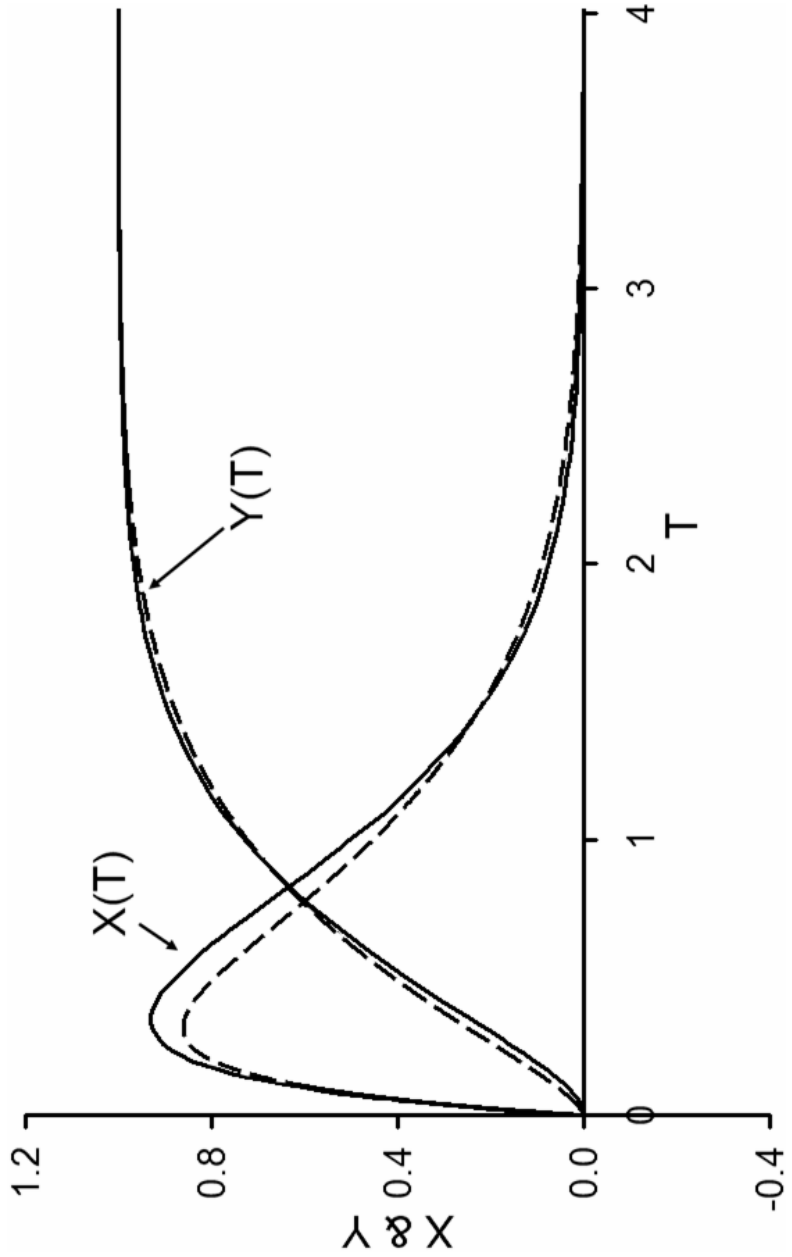

1.2
0.8
0.4
0.0
-0.4
X
0
1
2
3
4
T
X0(T)
Xδ(T)
Xε(T)
1.0
0.6
0.2
-0.2
-0.6
Y
0
1
2
3
4
T
Y0(T)
Yδ(T)
Yε(T)

Figure 2.2: Comparison of Solutions, Perturbation and Exact. These curves compare the composite perturbation solutions (dashed line) and exact numerical solutions (solid line) for $\gamma = 0.333$, $\delta = 0.111$, and $\varepsilon = 0.110$.


X(T)
Y(T)
T
X & Y
1.2
0.8
0.4
0.0
-0.4
0
1
2
3
4

scheme (A5) ensures that relative errors are within $O(\delta+\varepsilon)^2 \cong 0.05$ in this case.

Next we compare six different solutions of the MM equations – an exact numerical solution, our perturbation approximation, and three uniformly-valid approximations offered previously by others - for four different combinations of the MM parameters, chosen to explore regions of parameter space where different approximations are likely to be valid or to fail. We make these comparisons with phase plots of $C(t)$ vs. $S(t)$, since the phase-plot construction adequately reveals the quantitative differences while reducing the density of visual data by a factor of two.

We chose three other uniformly-valid approximate solutions for this comparison to represent the diversity of thought and effort that MM problem has inspired. Before presenting the plots we will review each approximation briefly and use the inequalities in Table 2.1 to delimit their strengths and weaknesses. We do not offer any of the more rudimentary QSSA solutions for comparison here, because they cannot meet one of the initial conditions for the MM problem, hence they generate irrational phase plots.

The first approximate perturbation solution of the MM equations we use for our comparison was presented by Heineken, Tsuchiya, and Aris (HTA) in 1967 **[124]**. It was the first treatment that went beyond a QSSA and included the full time-dependence of all the concentrations. They scaled the MM equations and solved them with the same singular perturbation method we have used, but, taking inspiration from the original sQSSA approximation **[114]**, they began with the *ad hoc* assumption that $\varepsilon = \frac{E_T}{S_T}$ was the correct singular perturbation parameter and adjusted the scaling of other terms in the equations to accommodate this choice. Their scaled equations include two dimensionless ratios – $\kappa = \frac{K_M}{S_T}$ and $\lambda = \frac{K}{S_T}$ – as coefficients of ostensibly $O(1)$ functions of the dimensionless concentration variables, and these

ratios appear as coefficients in the equations at all orders of the perturbation hierarchy. Consequently, HTA's perturbation solution should converge in an orderly way only when the perturbation parameter $\varepsilon \to 0$, of course, and when κ and λ are $O(1)$ or smaller. The criterion $\varepsilon \to 0$, or equivalently $E_T << S_T$, is met when $\mu \to 0$ according to (17), and the criteria $\kappa, \lambda \le O(1)$ are met for all circumstances except $K_M >> S_T$, or alternatively $\gamma \to 0$ according to (9). Based on these arguments, one anticipates that the HTA solution we present in Figures 2.3-2.6 will converge to the exact solution when μ is small but γ is not, a circumstance which is accounted for by rows (h-k) of Table 2.1.

The authors of two subsequent papers criticized HTA's scaling arguments and presented alternative perturbation solutions. First Seshadri and Fritzsch argued that $\varepsilon = \dfrac{E_T}{K_M}$ is the correct singular perturbation parameter **[123]**, and later Segel and Slemrod argued instead that $\varepsilon = \dfrac{E_T}{S_T + K_M}$ is correct **[117]**. These treatments evidence the same problem as HTA's, however, as they predict the perturbation functions will be $O(1)$ only if certain dimensionless coefficients that appear in the scaled equations are $O(1)$ or smaller. We have analyzed these solutions in thorough detail**,** and we were surprised to discover that both are in fact *identical* to the original HTA solution we use here for comparison. To verify this redundancy, one needs only to return all three analytical solutions from dimensionless to dimensioned, physical variables and compare them. Although each group of authors used different physical arguments to justify their scaling choices, in the end they achieve dominant balances among the same terms in the governing equations, which ensures that equivalent parts of the solutions appear at the same orders in their perturbation hierarchies. This unfortunate confusion and duplication of effort reveals how subtle the scaling and ordering exercise can be.

The second solution we use for comparison was derived by Schnell and Mendoza (SM) **[120]**, who recognized that the governing equations are fortuitously linear and admit an exact solution whenever $K_M >> S_T$, regardless of whether $E_T << S_T$ or vice versa. The ratio of short and long timescales in this limit is $\varepsilon = \frac{KE_T}{(E_T + K_M)^2}$. Since the only restriction for validity of their approach is $K_M >> S_T$, or $\gamma \to 0$ according to (9), the SM solution we present in Figures 2.3-2.6 is accurate for a wider range of circumstances than its authors claim, i.e. rows (a) through (g) of Table 2.1. We present the SM solution in the context of a different discussion as equation (A16) of the Appendix.

The third solution we use for comparison was derived by Tzafriri and Edelman (TE) **[116]** and was discussed in detail in a previous section. These authors correctly identified the short and long timescales; hence their solution is the first to give a reasonably accurate solution for all possible combinations of the MM parameters. Since the TE solution includes contributions at all orders from the equivalent of our regular perturbation parameter δ, but only the $O(1)$ contribution from the equivalent of our singular perturbation parameter ε, its estimated error relative to the exact solution is $O(\varepsilon)$, which is to be compared in Figures 2.3-2.6 with a relative error of $O(\delta + \varepsilon)^2$ for our perturbation solution.

Table 2.2 lists the MM parameters and associated dimensionless groups that were used to calculate the comparative phase plots of $C(t)$ vs. $S(t)$ in Figures 2.3 through 2.6.

Figure 2.3 compares the exact solution and four approximations for Case I of Table 2, which involves excess substrate $S_T >> E_T$ and the limit $\mu \to 0$ (=0.003), so traditional simplifications like the sQSSA and tQSSA are approximately valid and can be used to calculate $S(t)$ or $P(t)$ [but not $E(t)$ or $C(t)$]. This criterion also ensures $\delta, \varepsilon \to 0$ in our representation ($\delta = 0.002$, $\varepsilon = 0.002$), but the other parameters

$\gamma = 0.663$ and $\eta = \phi = 0.167$ are not at limiting values, so row (h) of Table 2.1 applies. As expected, the HTA, TE, and our solution all agree with each other and the exact solution in this case; indeed, they are indistinguishable within the resolution of the lines on the figure. The SM solution departs significantly, however, because $\gamma = 0.663$ is well outside its range of validity ($\gamma \to 0$).

Figure 2.4 compares them again for Case II of Table 2.2, which involves excess enzyme $E_T >> S_T$ and the limits $\gamma \to 0$ (=0.003) and $\delta \to 0$ (=0.002). The other parameters $\eta = \phi = 0.167$, $\mu = 0.663$, and $\varepsilon = 0.110$ are not at limiting values, so row (b) of Table 2.1 applies. This case demonstrates the strength of the SM solution, which has moved into its range of validity and nearly duplicates the exact solution, and the weakness of the HTA solution, which now fails quite obviously because its singular perturbation parameter is $\varepsilon = \dfrac{E_T}{S_T} = 199$. The TE solution does well with an estimated relative error of $O(\varepsilon) \cong 0.1$, but our solution does better, since the estimated error is only $O(\delta + \varepsilon)^2 \cong 0.01$ in this case.

Figure 2.5 compares them again for Case III of Table 2.2, where $E_T = S_T$, and the limits $\eta \to 0$ (=0.003) and $\varepsilon \to 0$ (=0.001) are realized. The other parameters $\gamma = \mu \cong \phi = 0.33$ and $\delta = 0.111$ are not at limiting values, so row (l) of Table 2.1 applies. These are exactly the criteria that justify the rQSSA. Neither the HTA nor the SM solution was derived for this particular circumstance, so each of these deviates significantly, whereas the TE solution is best of all and virtually indistinguishable from the exact solution. This is because the combination $\delta = 0.111$ and $\varepsilon = 0.001$ exposes its particular strength: the error estimate $O(\varepsilon) \cong 0.001$ in this favorable circumstance is an order of magnitude lower than the estimate $O(\delta + \varepsilon)^2 \cong 0.01$ for our solution.

Finally, Figure 2.6 compares them for Case IV of Table 2.2, where $\gamma \cong \eta \cong \mu \cong 0.33$, $\phi = 0.003$, and $\delta \cong \varepsilon = 0.11$, so no dimensionless parameter is at a

favorable limiting value, and no mathematical simplification like a QSSA applies. The SM solution fails as it did in Case I, because $\gamma = 0.333$ is outside its range of validity $(\gamma \to 0)$, and the TE solution agrees with the exact solution to within its anticipated error of $O(\varepsilon) \cong 0.1$. Since the combination $\delta + \varepsilon = 0.221$ is near its worst-case limit of $\frac{1}{4}$ for our solution, Case IV a stringent test, but the accuracy of our perturbation solution is still good, because the error is limited to $O(\delta + \varepsilon)^2 \cong 0.05$. Recall that good agreement was also evident in Figure 2.2, which compared scaled versions of the $C(t)$ and $P(t)$ curves to an exact solution for the same set of dimensionless parameters.

The HTA solution shows remarkable accuracy in Figure 2.6, even though its perturbation parameter is $\varepsilon = \frac{E_T}{S_T} = 1$ in this case. This success is fortuitous, however, and should not be mistaken as evidence that this solution is generally superior. All one needs to do is begin with the MM parameters for Case IV in Table 2.2 and recalculate the phase curves for decreasing values of k_2, which decreases K and K_M concomitantly. One finds that our solution, TE's, and SM's behave rationally and maintain their relative accuracy as k_2 decreases, whereas the HTA solution becomes progressively worse and diverges to nonsensical behavior when k_2 is less than about 30.

Based on these comparisons and others for a broad range of MM parameters that we do not present here, we conclude that the HTA solution **[124]** (along with the redundant solutions from Seshadri and Fritzsch **[123]** and Segel and Slemrod **[117]**) is the most restrictive to use. One can easily choose MM parameters, particularly combinations like $E_T >> S_T + K_M$ (or alternatively $\mu \to 1$ in our representation) that put it far outside of its intended range of validity and cause its perturbation series to diverge nonsensically. This is not the case for the others. The SM solution **[120]** gives rational if not accurate estimates of the $C(t)$- and $S(t)$-curves in any

Table 2.2: Representative Combinations of MM Parameters. These are the combinations of MM parameters (with units) and associated dimensionless groups that are used to calculate $C(t)$ vs. $S(t)$ in Figures 2.3-2.6. Case I involves excess substrate and $\delta, \varepsilon, \mu \to 0$ (*cf.* equation (17), row (h) of Table 2.1), so the sQSSA and tQSSA are approximately valid. Case II involves excess enzyme and $\delta, \gamma \to 0$ (*cf.* equation (9), row (b) of Table 2.1). Case III involves $\varepsilon, \eta \to 0$ (*cf.* equation (18), row (l) of Table 2.1), so the rQSSA is approximately valid. Case IV is an extreme test of the accuracy of the perturbation solution (A14), since only $\phi \to 0$ (*cf.* equation (19), row (n) of Table 2.1), and the perturbation parameters are very near their limiting values $\delta + \varepsilon < \frac{1}{4}$.

Quantity	Case I	Case II	Case III	Case IV
E_T (mol/vol)	1	199	100	100
$S_{T,}$ (mol/vol)	199	1	100	100
k_1 (vol/mol-time)	1	1	1	1
k_{-1} (time^{-1})	50	50	99	1
k_2 (time^{-1})	50	50	1	99
K (mol/vol)	50	50	1	99
K_D (mol/vol)	50	50	99	1
$K_M = K + K_D$ (mol/vol)	100	100	100	100
β (mol/vol), equation (7)	0.663	0.663	33.3	33.3
θ_C (time), equation (8)	0.003	0.003	0.003	0.003
θ_P (time), equation (13)	6	0.030	3	0.030
γ, equation (9)	0.663	0.003	0.333	0.333
η, equation (18a)	0.167	0.167	0.003	0.330
μ, equation (17)	0.003	0.663	0.333	0.333
ϕ, equation (18b)	0.167	0.167	0.330	0.003
δ, equation (10)	0.002	0.002	0.111	0.111
ε, equation (12)	0.001	0.110	0.001	0.110

Figure 2.3: Case I. These phase plots compare the HTA **[124]**, the SM **[120]**, the TE **[116]**, the perturbation solution described in this work, and an exact numerical solution for the set of MM parameters listed as Case I in Table 2.2. Only the SM solution fails to track the exact solution in this circumstance. The arrow points in the direction of increasing time.

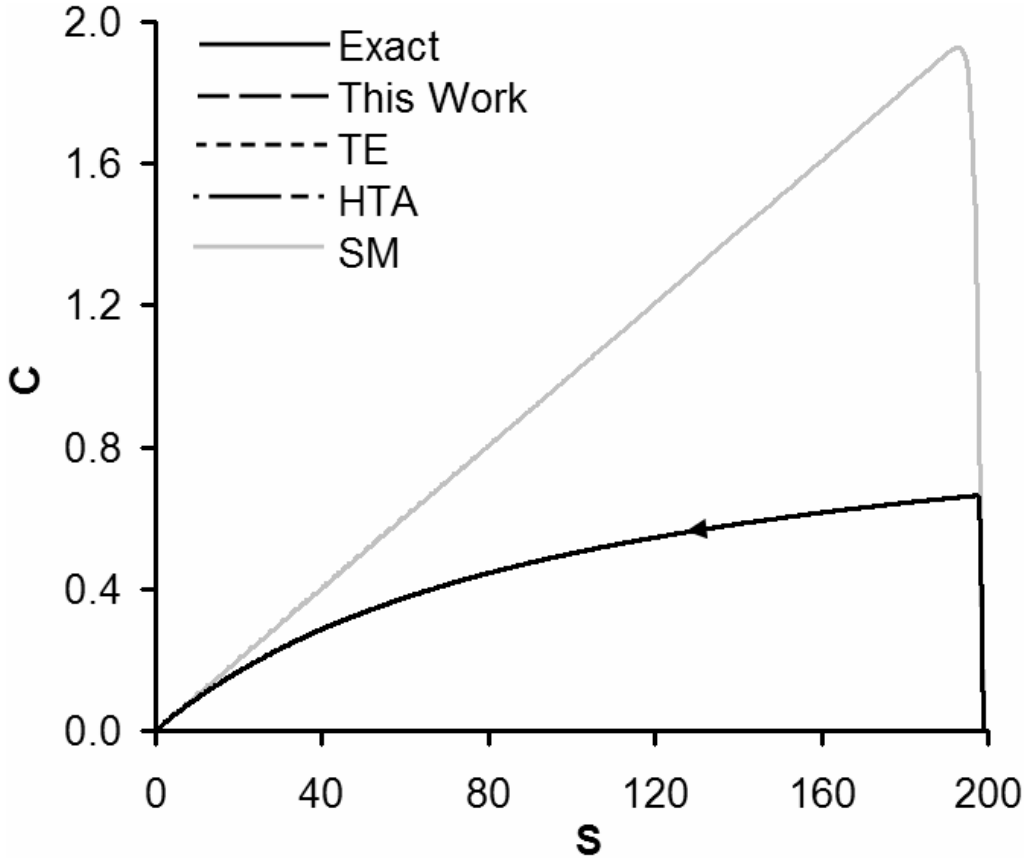

Exact
This Work
TE
HTA
SM
2.0
1.6
1.2
0.8
0.4
0.0
C
0
40
80
120
160
200
S

Figure 2.4: Case II. These phase plots compare the the HTA **[124]**, the SM **[120]**, the TE **[116]**, the perturbation solution described in this work, and an exact numerical solution for the set of MM parameters listed as Case II in Table 2.2. The SM solution is virtually identical to the exact solution in this case. The arrow points in the direction of increasing time.

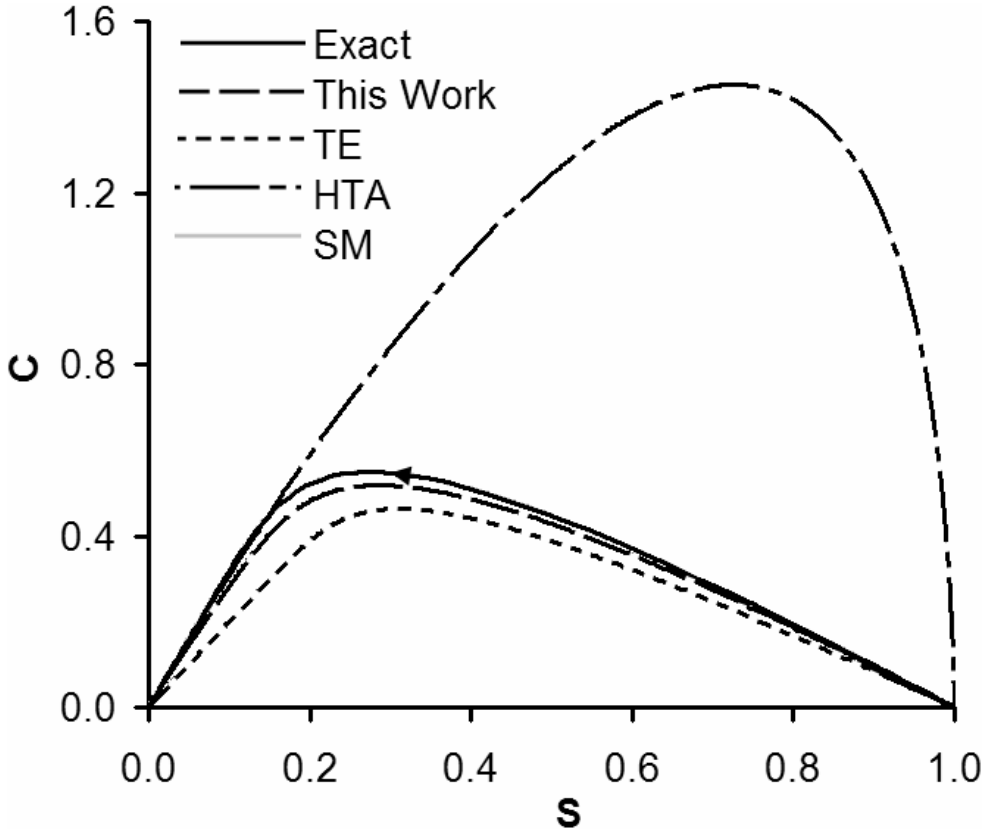

Exact
This Work
TE
HTA
SM
1.6
1.2
0.8
0.4
0.0
C
0.0
0.2
0.4
0.6
0.8
1.0
S

Figure 2.5: Case III. These phase plots compare the HTA **[124]**, the SM **[120]**, the TE **[116]**, the perturbation solution described in this work, and an exact numerical solution for the set of MM parameters listed as Case III in Table 2.2. The TE solution is virtually identical to the exact solution in this case. The arrow points in the direction of increasing time.

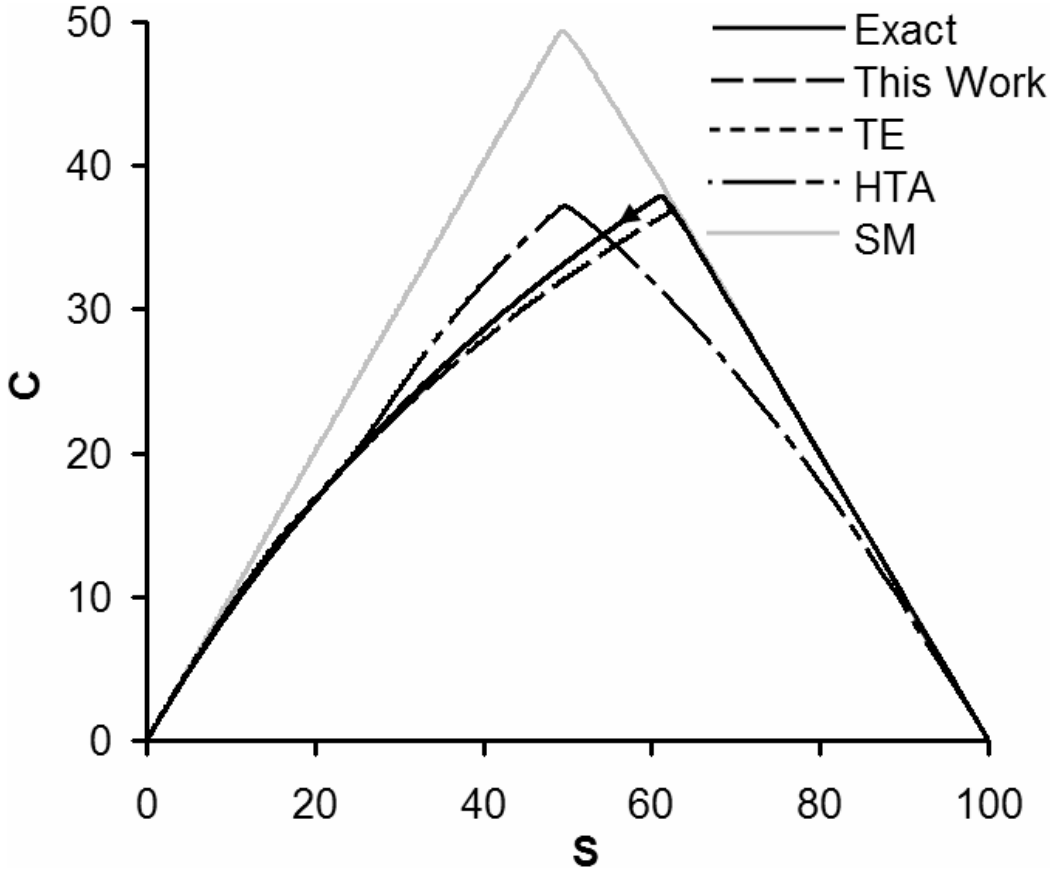
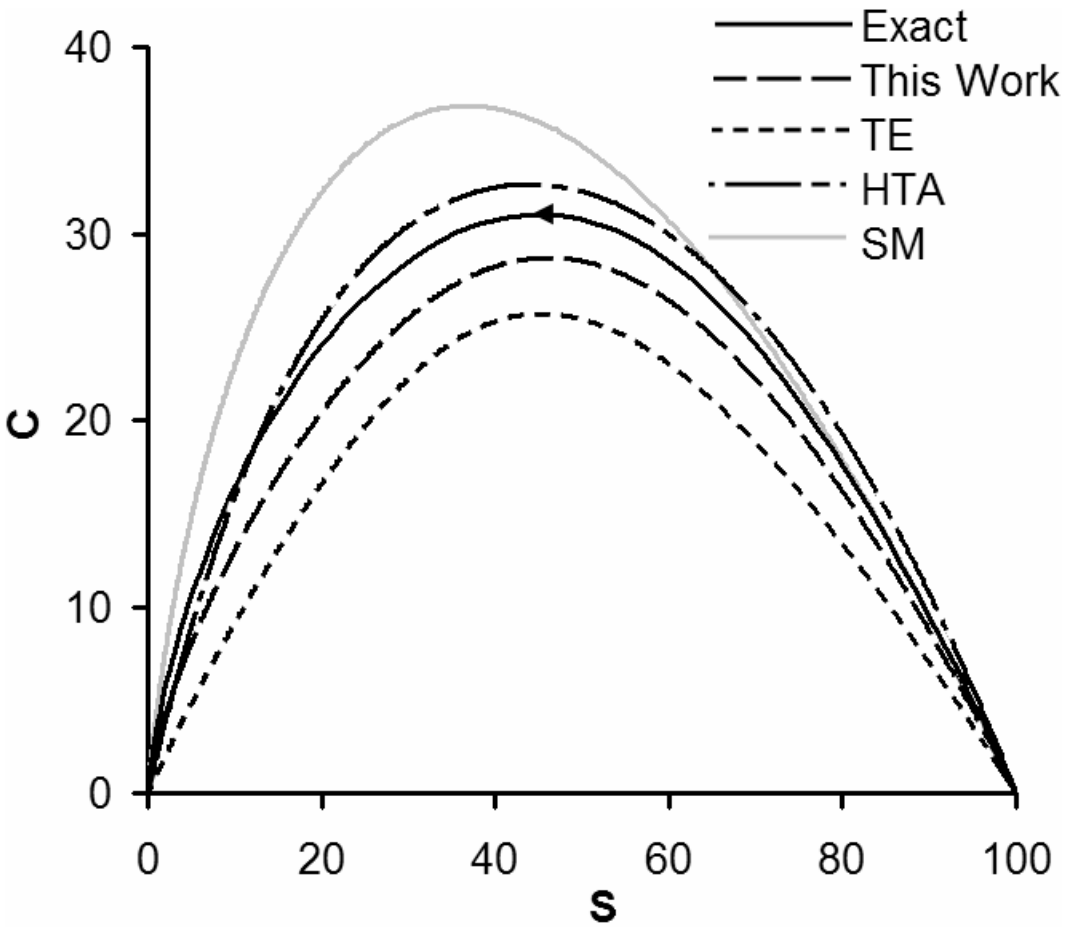

Exact
This Work
TE
HTA
SM
50
40
30
20
10
0
c
0
20
40
60
80
100
s

Figure 2.6: Case IV. These phase plots compare the HTA **[124]**, the SM **[120]**, the TE **[116]**, the perturbation solution described in this work, and an exact numerical solution for the set of MM parameters listed as Case IV in Table 2.2. The arrow points in the direction of increasing time.


Exact
This Work
TE
HTA
SM
40
30
20
10
0
C
0
20
40
60
80
100
S

circumstance, whereas the TE solution **[116]** and ours always give quantitatively accurate curves with errors limited only by the asymptotic errors of their missing, higher-order terms – $O(\varepsilon)$ for the former and $O(\delta+\varepsilon)^2$ for the latter. In most circumstances our solution is more accurate than the TE solution, because it includes an important $O(\varepsilon)$ contribution that the TE solution does not.

2.6 Conclusions

We have developed a new dimensionless representation of the MM equations, and we have found an approximate solution with a combined regular and singular perturbation scheme. The new scaling we introduce is the first that constrains all dimensionless concentration variables to be $O(1)$ for all time, which ensures that our solution converges uniformly and accurately for any physically realizable combination of MM parameters. As a by-product of our scaling analysis we have identified four dimensionless ratios – γ, η, μ, and ϕ (*cf.* Table 2.1) – that serve to categorize all the previous approximations for the MM problem (sQSSA, integrated MM equation, initial-rate method, tQSSA, rQSSA, singular perturbations, and others), to delimit their accuracy, and to reveal some new regions of parameter space where traditional QSSAs for the MM problem are valid. Our perturbation solution properly shows the leading-order effects of an unavoidable quadratic nonlinearity in the MM equations (*cf.* δ, equation (10)), and it includes important leading-order corrections for the effects of two widely disparate timescales in the MM equations (*cf.* ε, equation (12)). The cumulative effect of our analysis is a comprehensive, unprecedented description of the mechanistic subtleties of this important problem.

The only obvious avenue for progress is to improve accuracy by finding the neglected, higher-order terms our perturbation solution has omitted. One can envisage

two ways to proceed. The first (and easiest) method would be to extend the perturbation expansion, equation (A5), to include terms of $O(\delta^2)$, $O(\varepsilon^2)$, and $O(\delta\varepsilon)$, and to solve the higher-order equations, which would reduce the maximum relative error in the composite solution to $O(\delta+\varepsilon)^3 = \left(\frac{1}{4}\right)^3 \cong 0.02$. The higher-order equations are certain to be linear, hence soluble by the same methods we used to find the $O(\delta)$ and $O(\varepsilon)$ terms, but the effort would be formidable, and the reward would be small. A second (and more difficult) way to proceed would be to use the TE solution **[116]**, which already has higher-order effects of δ in it, as the $O(1)$ starting point for a singular perturbation in ε alone. However, the effort here would be even more intimidating, since the TE solution is the product of two solutions of nonlinear equations and is significantly more complicated than ours. Perhaps the only reasonable thing to consider is to "patch" the gap in the TE solution by adding our $O(\varepsilon)$ correction to it, but even then, the remaining errors would still be $O(\varepsilon)^2 = \frac{1}{16}$. We conclude there is little incentive to proceed.

Finally, we offer a few comments on the pedagogical value of this work. The scaling and perturbation methods we have used here are standard tools of engineering analysis. Many of us are charged with the duty of teaching others how to use these tools properly, and we seek example-problems that combine practical relevance with the potential for "clean" solutions that our students can easily derive and understand. Generations of chemical engineering graduate students have "cut their teeth" on singular perturbation theory by repeating Heineken, Tsuchiya, and Aris's classic analysis of this important MM problem **[124]**; indeed, it has been reproduced in graduate-level textbooks as a prototype for the singular perturbation method **[91, 127, 128]**. We believe our new analysis of the MM problem is actually a better example to learn from, because our scaling more transparently and effectively accomplishes the general goals of scaling analysis; our scaled perturbation equations (A6) and (A8) are

relatively easy to solve; and our method introduces regular and singular perturbations together in the context of solving a single problem, which helps beginners recognize and appreciate their differences.

2.7 Acknowledgements

This project was supported by the Initiative for Future Agriculture and Food Systems Grant no. 2001-52104-11484 from the USDA Cooperative State Research, Education, and Extension Service. We thank our colleague Paul Steen for helpful comments.

2.8 Appendix – Perturbation Solution of Equations (14-15)

We begin by solving the inner equations (14). Our first move is to postulate a solution in the form of a perturbation expansion in ε as follows:

(A1a) $x(\tau) = x_0(\tau) + \varepsilon x_\varepsilon(\tau) + O(\varepsilon^2)$;

(A1b) $y(\tau) = y_0(\tau) + \varepsilon y_\varepsilon(\tau) + O(\varepsilon^2)$;

where, x_0, x_ε, y_0 and y_ε are as-yet unknown functions of ε. These functions must meet the initial conditions of the original problem for arbitrary values of ε, in which case $x_0(0) = y_0(0) = x_\varepsilon(0) = y_\varepsilon(0) = 0$. Substituting the (A1) into (14) and collecting terms by orders in ε, one finds at $O(1)$:

(A2a) $\frac{dx_0}{d\tau} = 1 - x_0 + \delta x_0^2 - (1 - \gamma x_0) y_0$;

(A2b) $\frac{dy_0}{d\tau} = 0$;

and at $O(\varepsilon)$:

(A3a) $\frac{dx_\varepsilon}{d\tau} = (-1 + 2\delta x_0 - y_0)x_\varepsilon - (1 - \gamma x_0)y_\varepsilon$;

(A3b) $\frac{dy_\varepsilon}{d\tau} = x_0(\tau)$.

Neglecting terms of $O(\varepsilon^2)$ and higher introduces errors $O\left(\frac{1}{16}\right)$ or smaller, since $0 < \varepsilon < \frac{1}{4}$ according to equation (12).

From (A2b) one obtains the trivial solution $y_0 = 0$, and from (A2a), which is a nonlinear but soluble Riccati equation, one obtains

(A4) $$x_0(\tau) = \frac{1 - \exp\left[-\left(\sqrt{1-4\delta}\right)\tau\right]}{1 - \frac{1}{2}\left(1 - \sqrt{1-4\delta}\right)\left\{1 + \exp\left[-\left(\sqrt{1-4\delta}\right)\tau\right]\right\}}.$$

One finds quickly that the Riccati function makes equations (A3) for the $O(\varepsilon)$ corrections very tedious to solve. Recognizing this difficulty, and also recalling that $0 < \delta < \frac{1}{4}$ according to equation (10), we linearize equation (A2) by introducing *regular* perturbation expansions in the parameter δ for the functions $x(\tau)$ and $y(\tau)$. Now these functions become two-parameter expansions of the form

(A5a) $x(\tau) = x_0(\tau) + \delta x_\delta(\tau; \varepsilon = 0) + \varepsilon x_\varepsilon(\tau; \delta = 0) + O(\delta + \varepsilon)^2$;

(A5b) $y(\tau) = y_0(\tau) + \delta y_\delta(\tau; \varepsilon = 0) + \varepsilon y_\varepsilon(\tau; \delta = 0) + O(\delta + \varepsilon)^2$.

Since $\delta + \varepsilon < \frac{1}{4}$, including only the first term in δ and ε is sufficient to ensure that errors in $x(\tau)$ and $y(\tau)$, as quantified by the first neglected terms in the perturbation expansions, are $O(\delta + \varepsilon)^2$, which is once again $\frac{1}{16}$ or smaller.

One might wonder at this point why it is necessary to solve the equations with a two-parameter perturbation expansion. Since $\delta < \frac{1}{4}$, $\varepsilon < \frac{1}{4}$, and $\delta + \varepsilon < \frac{1}{4}$, why not recast the original equations (14) and (15) in terms of ε only and solve them with a one-parameter expansion, as others have in the past **[117, 123, 124]**? The reason is subtle and worthy of explanation. Note that the ratio $\frac{\delta}{\varepsilon} = \frac{S_T}{K} = \frac{\gamma}{\eta}$ [*cf.* equations (9), (10), (12), and (18)] can range anywhere between zero and infinity, so δ and ε are in general of different order, even though they are independently bounded. If one solves the problem with a one-term singular perturbation expansion in ε, the unbounded ratio $\frac{\gamma}{\eta}$ appears unavoidably as a coefficient in the $O(\varepsilon)$ equations; consequently, the $O(\varepsilon)$ functions in the perturbation hierarchy are $O\left(\frac{\gamma}{\eta}\right)$ rather than $O(1)$ and can diverge when this ratio is large. Using a two-term perturbation expansion in δ and ε circumvents this difficulty by ensuring all terms are properly bounded.

Inserting (A5) into (14) and collecting terms by orders in δ and ε, one finds the following six equations to solve on the inner timescale τ.

(A6a) $\frac{dx_0}{d\tau} = 1 - x_0 - y_0 + \gamma x_0 y_0$ with $x_0(0) = 0$;

(A6b) $\frac{dy_0}{d\tau} = 0$ with $y_0(0) = 0$;

(A6c) $\frac{dx_\delta}{d\tau} = x_0^2 - x_\delta - y_\delta + \gamma x_\delta y_0 + \gamma x_0 y_\delta$ with $x_\delta(0) = 0$;

(A6d) $\frac{dy_\delta}{d\tau} = 0$ with $y_\delta(0) = 0$;

(A6e) $\dfrac{dx_\varepsilon}{d\tau} = -x_\varepsilon - y_\varepsilon + \gamma x_\varepsilon y_0 + \gamma x_0 y_\varepsilon$ with $x_\varepsilon(0) = 0$;

(A6f) $\dfrac{dy_\varepsilon}{d\tau} = x_0$ with $x_\varepsilon(0) = 0$.

The solution of these is straightforward, as they are all linear and coupled only in a cascading way. After solving at each order and substituting into (A5), one obtains the inner solutions:

(A7a) $$\begin{aligned} x(\tau) &= x_0 + \delta x_\delta + \varepsilon x_\varepsilon = 1 - e^{-\tau} + \delta\left[1 - \left(2\tau + e^{-\tau}\right)e^{-\tau}\right]\cdots \\ &+ \varepsilon\left\{(\gamma - 1)(\tau - 2) + \left[(2\gamma - 1)\tau - \frac{1}{2}\gamma\tau^2 - 2 + \gamma\right]e^{-\tau} + \gamma e^{-2\tau}\right\}; \end{aligned}$$

(A7b) $y(\tau) = y_0 + \delta y_\delta + \varepsilon y_\varepsilon = \varepsilon\left(e^{-\tau} + \tau - 1\right).$

Note that if one expands the exact Riccati solution (A4) for small δ, one obtains exactly the $O(1)$ and $O(\delta)$ portions of (A7a), above, which confirms that the perturbation in δ linearizes equation (A2a).

We solve the outer equations (15) by inserting perturbation series for $X(T)$ and $Y(T)$ analogous to equations (A1), above, and collecting terms by order in δ and ε, which generates the following set of equations to solve on the outer timescale T:

(A8a) $0 = 1 - X_o - Y_o + \gamma X_o Y_o$;

(A8b) $\dfrac{dY_o}{dT} = X_o$;

(A8c) $0 = X_o^{\,2} - X_\delta - Y_\delta + \gamma Y_o X_\delta + \gamma X_o Y_\delta$;

(A8d) $\frac{dY_\delta}{dT} = X_\delta$;

(A8e) $\frac{dX_\varepsilon}{dT} = -X_\varepsilon - Y_\varepsilon + \gamma Y_o X_\varepsilon + \gamma X_o Y_\varepsilon$;

(A8f) $\frac{dY_\varepsilon}{dT} = X_\varepsilon$.

Since two derivatives have been lost in the ordering scheme (*cf.*, A8a and c), the solutions of these equations cannot meet all the initial conditions of the overall problem, so one must solve them in terms of integration constants that will be determined by matching the outer solutions asymptotically to the inner solutions. Rearranging (A8a), one finds

(A9a) $X_o(Y_0) = \frac{1 - Y_o}{1 - \gamma Y_o}$;

which, when substituted into(A8b) and integrated gives

(A9b) $T(Y_o) = \gamma Y_o + (\gamma - 1)\ln(1 - Y_o) + A$;

where A is an integration constant. Equation (A9b) gives $Y_0(T)$ implicitly rather than explicitly, which proves inconvenient when solving the remaining outer equations, as they depend explicitly on $Y_0(T)$. To circumvent this difficulty, we use (A8b) to transform the independent variable from T to Y_0 in the remaining equations, i.e., $\frac{d}{dT} = \left(\frac{dY_0}{dT}\right)\frac{d}{dY_0} = X_0(Y_0)\frac{d}{dY_0}$. The final outer solutions in terms of Y_0 are

$$X(Y_o) = X_0 + \delta X_\delta + \varepsilon X_\varepsilon = \frac{1-Y_o}{1-\gamma Y_o}\cdots$$

$$+\delta\left\{\frac{(1-Y_o)^2}{(1-\gamma Y_o)^3}\left[1-\left(\frac{1-\gamma}{1-Y_o}\right)\left(B-\frac{1}{\gamma}\ln(1-\gamma Y_o)\right)\right]\right\}\cdots \quad ; \tag{A10a}$$

$$+\varepsilon\left\{\left(\frac{1-Y_o}{(1-\gamma Y_o)^3}\right)\left[(1-\gamma)\left(\frac{2}{1-\gamma Y_o}-C\right)+\ln\left(\frac{1-Y_o}{1-\gamma Y_o}\right)\right]\right\}$$

$$Y(Y_o) = Y_0 + \delta Y_\delta + \varepsilon Y_\varepsilon = Y_o + \delta\left[\frac{(1-Y_o)}{(1-\gamma Y_o)}\left(B-\frac{1}{\gamma}\ln(1-\gamma Y_o)\right)\right]\cdots$$

$$+\varepsilon\left\{\left(\frac{1-Y_o}{1-\gamma Y_o}\right)\left[C+\frac{1}{\gamma Y_o-1}+\frac{1}{\gamma-1}\ln\left(\frac{1-Y_o}{1-\gamma Y_o}\right)\right]\right\} \quad ; \tag{A10b}$$

where B and C are integration constants, and $Y_0(T)$ is given implicitly by equation (A9b).

We must determine the integration constants A, B, and C by asymptotically matching the inner and outer solutions in the overlap region, where $\tau >> 1$ and $T << 1$. The so-called "matching solutions" can be identified by recasting the inner solutions in terms of the outer variable T and finding their asymptotic behavior in the limit $\varepsilon \to 0$ with T = finite. Once these are available, the integration constants in the outer solutions can be identified by recasting the outer solutions in terms of the inner variable τ, finding their asymptotic behavior in the limit $\varepsilon \to 0$ with τ = finite, and requiring them to reproduce the matching solutions. Composite solutions that are uniformly valid for the entire range $0 < t < \infty$ are constructed by summing the inner and outer solutions and subtracting away the redundant matching solution for the overlap region. This procedure is outlined in any of the numerous textbooks on singular perturbation theory **[91, 102, 103]**. One has in our case the added complication that the inner and outer solutions must match for arbitrary values of the regular perturbation parameter δ, but this introduces little difficulty.

The matching solutions derived from the inner solutions are

(A11a) $x_{match}(T) = 1 + (\gamma - 1)T + \delta - 2\varepsilon(\gamma - 1)$;

(A11b) $y_{match}(T) = T - \varepsilon$.

Finding the outer matching solutions is complicated by the fact that the outer solutions are given explicitly in terms of Y_0 rather than T. Since $Y_0 \to 0$ when $T \to 0$, we expand (A9b) for small Y_0, which gives

(A12a) $T \approx A + Y_o + \frac{1}{2}(1-\gamma)Y_o^{\,2} + O\left(Y_o^{\,3}\right)$;

and we invert this expansion to obtain

(A12b) $Y_0 \approx (T - A) + \frac{1}{2}(\gamma - 1)(T - A)^2 + O(T - A)^3$.

Comparing to (A11a) requires $A = 0$ for matching at $O(1)$. Substituting the resulting expansion into the other outer solutions and expanding them for small T, one finds

(A13a) $X(T) \approx 1 + (\gamma - 1)T + O(T^2) + \delta[1 + B(\gamma - 1) + O(T)] \cdots$
$+ \varepsilon[(C - 2)(\gamma - 1) + O(T)]$;

(A13b) $Y(T) \approx T + O(T^2) + \delta[B + O(T)] + \varepsilon[C - 1 + O(T)]$.

Comparing to equations (A11) reveals $B = C = 0$.

One can now combine all of the parts, i.e. (inner, A7) + (outer, A10) – (matching, A11), to formulate the uniformly valid composite solutions. We present

them here in terms of the outer independent variable T (=$\varepsilon\tau$), since it is the relevant timescale for product formation.

(A14a) $$T = \gamma Y_o - (1-\gamma)\ln(1-Y_o);$$

(A14b)
$$\begin{aligned} X(T) &= X_0 + \delta X_\delta + \varepsilon X_\varepsilon + O(\delta+\varepsilon)^2 = \frac{1-Y_o}{1-\gamma Y_o} - e^{-\frac{T}{\varepsilon}} \cdots \\ &+ \delta\left[\frac{(1-Y_o)^2}{(1-\gamma Y_o)^3} + \frac{(1-\gamma)(1-Y_o)}{\gamma(1-\gamma Y_o)^3}\ln(1-\gamma Y_o)\right] - \delta e^{-\frac{T}{\varepsilon}}\left[2\left(\frac{T}{\varepsilon}\right) + e^{-\frac{T}{\varepsilon}}\right]\cdots \\ &+ \varepsilon\frac{1-Y_o}{(1-\gamma Y_o)^3}\left[\frac{2(1-\gamma)}{1-\gamma Y_o} + \ln\left(\frac{1-Y_o}{1-\gamma Y_o}\right)\right]\cdots \\ &- \varepsilon e^{-\frac{T}{\varepsilon}}\left[\frac{\gamma}{2}\left(\frac{T}{\varepsilon}\right)^2 + (1-2\gamma)\left(\frac{T}{\varepsilon}\right) + 2 - \gamma\left(1+e^{-\frac{T}{\varepsilon}}\right)\right] + O(\delta+\varepsilon)^2 \end{aligned};$$

(A14c)
$$\begin{aligned} Y(T) &= Y_0 + \delta Y_\delta + \varepsilon Y_\varepsilon + O(\delta+\varepsilon)^2 = Y_o - \delta\frac{1-Y_o}{\gamma(1-\gamma Y_o)}\ln(1-\gamma Y_o)\cdots \\ &- \varepsilon\left\{\left(\frac{1-Y_o}{1-\gamma Y_o}\right)\left[\frac{1}{1-\gamma Y_o} + \frac{1}{1-\gamma}\ln\left(\frac{1-Y_o}{1-\gamma Y_o}\right)\right] - e^{-\frac{T}{\varepsilon}}\right\} + O(\delta+\varepsilon)^2 \end{aligned}.$$

One can invert equation (A14a) by choosing values of Y_0 ranging from 0 to 1 and calculating corresponding values of T, which are then used to solve equations (A14b-c) for X and Y.

Given values of the initial concentrations E_T and S_T and the kinetics parameters k_1, k_{-1}, and k_2 (or alternatively K, K_D, and K_M), one can calculate values of the various dimensionless groups and scaling parameters in equations (7-13), and then use equations (A14) to find $C = \beta X$ and $P = S_T Y$. Finally, $E(t)$ and $S(t)$ can be obtained if necessary from the mass constraints (4). This procedure was used to construct Figures 2.3-2.6.

The effects of the perturbation parameters δ and ε in (A14) are obvious, but the effects of γ are not. Recall that $\gamma = \dfrac{\beta}{E_T}$ is the fraction of enzyme initially charged to the reactor that forms complex on the short timescale, so $0 < \gamma < 1$, and that $\gamma \to 0$ ensures the free enzyme concentration remains nearly constant [*cf.* equation (9), and Table 2.1, rows (a)-(g)]. If one is interested specifically in the limiting circumstance $\gamma \to 0$, it makes sense to reexamine the original problem, equations (14-15), treating γ as the regular perturbation parameter instead of δ, because $\delta = \gamma\mu$ according to equation (20), and $0 < \mu < 1$ according to (17). The procedure in this case parallels exactly what we outlined in (A5) through (A14). One finds:

$$X(T) = e^{-T} - e^{-\frac{T}{\varepsilon}} + \gamma\left\{(1-T)e^{-T} + (\mu-1)e^{-2T} - \mu\left[2\left(\frac{T}{\varepsilon}\right) + e^{-\frac{T}{\varepsilon}}\right]e^{-\frac{T}{\varepsilon}}\right\}\cdots$$
$$+\varepsilon\left[(2-T)e^{-T} - \left(2+\frac{T}{\varepsilon}\right)e^{-\frac{T}{\varepsilon}}\right] + O(\varepsilon+\gamma)^2 \quad ; \tag{A15a}$$

$$Y(T) = 1 - e^{-T} + \gamma T e^{-T} + \varepsilon\left[e^{-\frac{T}{\varepsilon}} - (1-T)e^{-T}\right] + O(\varepsilon+\gamma)^2 . \tag{A15b}$$

This is effectively the small-γ limit of our more complete perturbation solution (A14), above, which is difficult to obtain by manipulating (A14) directly. When $\gamma = 0$ exactly, $\delta = 0$ also, and the governing equations (14) or (15) are fortuitously linear, thus they can be solved exactly for arbitrary ε. This circumstance was first described and an analytical solution given by Schnell and Mendoza **[120],** and we include it in our comparison of solutions, Figures 2.3-2.6. In terms of our scaled variables, one finds

(A16a) $$X(T)=\frac{1}{\sqrt{1-4\varepsilon}}\left\{\exp\left[-\left(\frac{1-\sqrt{1-4\varepsilon}}{2\varepsilon}\right)T\right]-\exp\left[-\left(\frac{1+\sqrt{1-4\varepsilon}}{2\varepsilon}\right)T\right]\right\};$$

(A16b) $$Y(T)=1+\left(\frac{1-\sqrt{1-4\varepsilon}}{2\sqrt{1-4\varepsilon}}\right)\exp\left[-\left(\frac{1+\sqrt{1-4\varepsilon}}{2\varepsilon}\right)T\right]\cdots -\left(\frac{1+\sqrt{1-4\varepsilon}}{2\sqrt{1-4\varepsilon}}\right)\exp\left[-\left(\frac{1-\sqrt{1-4\varepsilon}}{2\varepsilon}\right)T\right].$$

Since this solution includes contributions from all orders of ε, it is more accurate than our perturbation solution for this one restrictive case, $\gamma=\delta=0$. One recovers the $\gamma=\delta=0$ version of our perturbation solution by expanding (A16) to first order in ε.

There is only one circumstance for which $\gamma\to 1$, and that is $S_T >> E_T + K_M$ [*cf.* Table 2.1, row (i)]. In this unusual situation all enzyme binds quickly to a large excess of substrate, remaining bound and making product at a constant rate until the substrate is nearly consumed. The scaled complex curve $X(T)$ rises almost instantly from zero to one, stays there until $T\cong 1$, and then falls almost instantly back to zero. The scaled product curve $Y(T)$ rises linearly from zero to one at $T\cong 1$, turns sharply, and remains at one thereafter. This odd behavior for $\gamma\to 1$ is evidenced in the governing equations as an intriguing boundary singularity near $T=1$, and it requires that all terms in (A14a) for $Y_0(T)$ be retained in the limit $(1-\gamma)\to 0$. We attempted to solve the original equations (14) and (15) with a singular perturbation in the small parameter $(1-\gamma)$, which would serve to identify the leading-order effect of $\gamma\to 1$, but we were unsuccessful. This is of no practical consequence, however, because the full solutions (A14) converge uniformly for any physically realizable circumstance, i.e. $0<\gamma<1$.

CHAPTER 3

THE KINETICS OF P-NITROPHENYL-β-D-CELLOBIOSIDE HYDROLYSIS AND TRANSGLYCOSYLATION BY THERMOBIFIDA FUSCA CEL5ACD

3.1 Introduction

Glycosides of mono- and dinitrophenols are commonly used for studying glycosyl hydrolases because their hydrolysis yields an easily detectable nitrophenolic product **[92-94]**. These substrates typically contain mono- or di-saccharide glycones specific to the glycosyl hydrolase being studied and a nitrophenolic aglycone that mimics a sugar unit, insofar as it fits in the active site and does not interfere with the underlying hydrolytic mechanism. The small size of the substrate ensures high if not complete specificity for bond cleavage at the aglyconic bond where the nitrophenolic "reporter" group is attached. The enzymatic reactions can be terminated in alkaline conditions, deprotonizing a catalytic proton donor and increasing the absorbance of the reporter molecule, which usually has a peak around 400 nm and can be quantified with either an extinction coefficient or a standard curve.

Retaining glycosyl hydrolases add an additional layer of complexity, because their double-displacement mechanism contains a glycosyl-enzyme intermediate, which, if attacked by a carbohydrate nucleophile (often another substrate molecule), can result in transglycosylation **[42, 43, 131]**. In some cases increasing the substrate concentration reduces the reporter yield and leaves some reporter hidden in a non-hydrolyzable glycosidic product. When this happens, the reaction appears to be substrate-inhibited **[132, 133]**. Transglycosylation is fundamentally different from substrate inhibition, however, because in transglycosylation substrate is being

consumed in an alternate pathway, and in inhibition substrate is simply quenching the activity of the enzyme **[134, 135]**. The presence of an alternate pathway for transglycosylation complicates the mechanistic picture and the mathematical description of reaction kinetics for retaining glycosyl hydrolases and their mutants.

In this study we characterize the hydrolysis and transglycosylation of a glycoside of cellobiose and p-nitrophenol (pNP-G2) by the catalytic domain of *Thermobida fusca* Cel5A (Cel5Acd), a 34.5 kDa, endo-specific, family 5-2 glycosyl hydrolase **[26]**. We present a two-pathway mechanism for pNP-G2 hydrolysis by Cel5Acd that includes the formation of pNP-reporter and a non-hydrolyzable pNP-G1 product. The same reaction network and rate expression has been identified previously and used to fit to initial-rate data for reporter formation in other transglycosylating systems **[132, 133, 136]**. We advance the knowledge gained in the previous investigations by deriving rigorous conditions that delimit applicability of the rate expression and by integrating it predict the variation of reporter concentration for all times. We verify the accuracy of the integral solution and by inference the reaction mechanism through quantitative comparisons to pNP-vs.-time data collected for varying enzyme and substrate concentrations.

3.2 Experimental

Cel5Acd Production. Cel5Acd was purified from *E. coli* strain D1430 as described previously for a *T. fusca* xyloglucanase **[137]**, but with the following modifications. The P-sepharose column was eluted with successive high-salt to low-salt washes, using approximately 2-column-volumes of 1.2 M, 0.6 M, 0.3 M, and 0.0 M ammonium sulfate with a continuous 5 mM Kpi buffer (pH 6.0), with Cel5Acd eluting in the final wash. The Q-Sepharose column was eluted with a 6-column-

volume gradient from 0 to 0.5 M NaCl, with a continuous 10 mM BisTris buffer (pH 5.6). Final protein fractions were exchanged into sodium acetate buffer (50 mM NaAc, 0.02 % NaAz, pH 5.5) and concentrated to roughly 60 μM using a stirred cell ultrafiltration chamber with a polyethersulfone ultrafiltration membrane (10 kDa MWCO, Millipore, Billerica, MA).

Strain D1430 contains an internal L175Q mutation (Brian Barr, Personal Communication, Loyola University, Baltimore, MD). The mutation does not appear to affect the protein fold or activity, as it purified like wild-type and had comparable activity on homogenous substrates.

Hydrolysis of pNP-G2 by Cel5Acd. pNP-G2 (N5759, >98 % purity) and pNP (1048, spectrophotometric grade) were obtained from Sigma-Aldrich (St. Louise, MO). Solutions of pNP and pNP-G2 were prepared at 10 mM in sodium acetate buffer (50 mM NaAc, 0.02 % NaAz, pH 5.5) and stored at -20 ^{o}C.

All reactions were prepared in sodium acetate buffer (50 mM NaAc, 0.02 % NaAz, pH 5.5) from the same initial enzyme and substrate solutions. Activity assays of Cel5Acd suspended in sodium acetate buffer for up to 6 days at 50 ^{o}C confirmed that the enzyme remained stable for the entire duration of the reaction under the conditions employed. For each product curve Cel5Acd, pNP-G2, and buffer were combined in a 10 mL screw-cap plastic tube, shaken, and transferred in 100 μL aliquots into separate test tubes. The tubes were covered and placed in a 50 ^{o}C water bath. Every 15 minutes three tubes were selected at random and moved to a cool water bath, where 1 mL of 2 M Na_2CO_3 was added to stop the reaction. After the last time-point (195 minutes), 2 mL of deionized (DI) water were added to each tube, and the optical density was measured with a spectrophotometer at 400 nm (OD400). The OD400 readings were adjusted using the average zero-time value as a blank, and pNP concentrations were calculated by comparing to a standard curve.

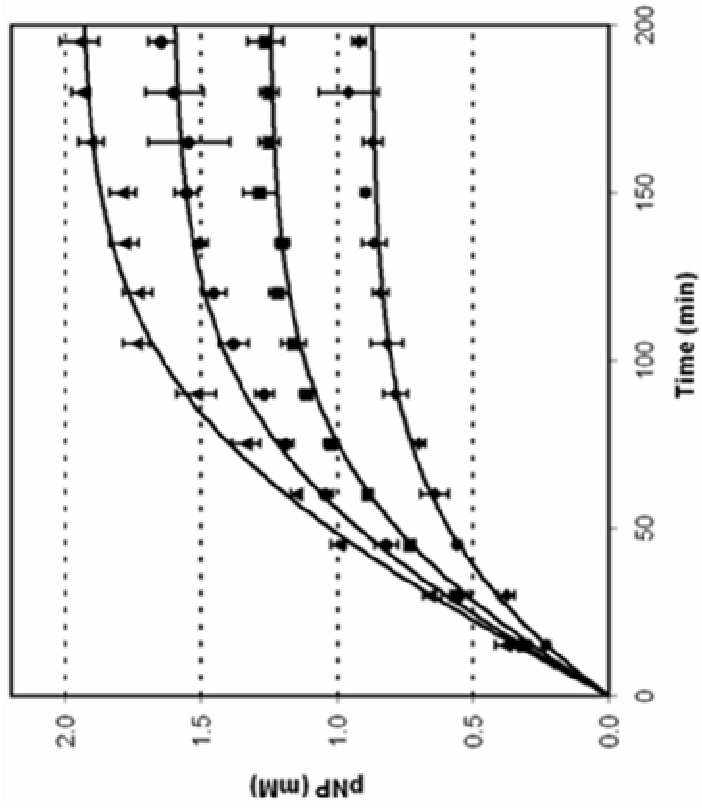
Hydrolysis reactions for thin layer chromatography (TLC) were run in 100 μL volumes as described above but without the addition of Na_2CO_2 or DI water. At completion, 10 μL aliquots were spotted directly onto a TLC plate (Whatman LK5D, 150 Å silica gel plates), dried, separated, and developed as described previously **[138-140]**.

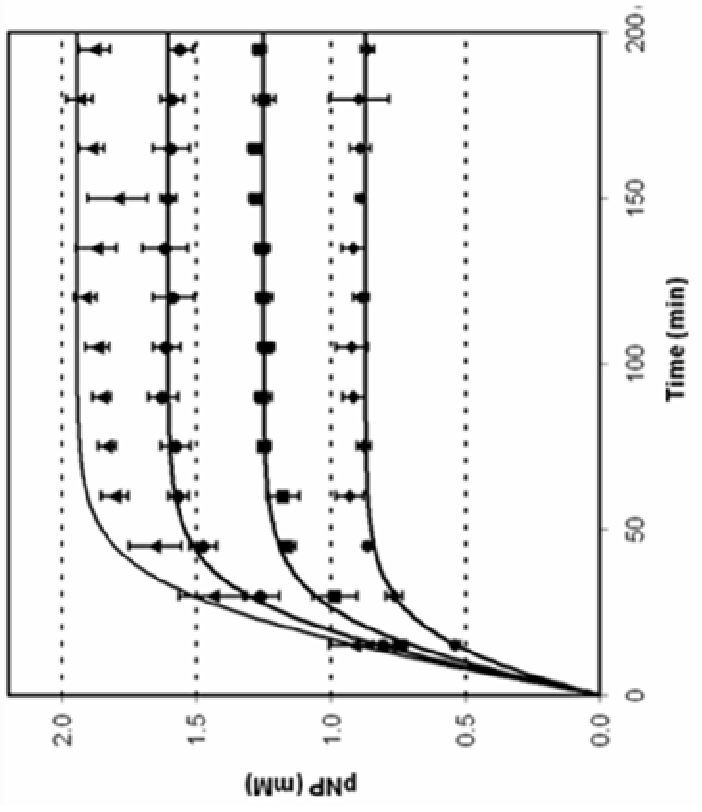
Data analysis. pNP concentrations were measured at 13 equal time increments of 15 min each for all combinations of two different enzyme concentrations (1.4 and 4.0 μM) and four different substrate concentrations (1.0, 1.5, 2.0, and 2.5 mM). Each measurement was repeated three times to generate an average pNP concentration with an associated (sample) standard deviation. All 104 data points were fit simultaneously with a four-parameter, model-based analytic function, which is described in section 3.4, below. Best-fit values of the four adjustable parameters were determined by minimizing χ^2, the weighted sum of squared errors, via a Monte Carlo search of parameter space **[141]**. The standard deviation of each fit-parameter was estimated by forcing the parameter to vary from its best-fit value, re-minimizing χ^2 with respect to the other three parameters, and using the results to estimate the curvature of the χ^2 function **[141]**. Because the other parameters are allowed to vary in this calculation, the standard deviations of fit parameters grossly include the effects of any correlations among them, although they are not explicitly calculated here **[5]**. The minimization program was written in-house using Visual Basic/Excel.

3.3 Results

pNP-G2 hydrolysis by Cel5Acd. The pNP product curves for eight combinations of initial Cel5Acd and pNP-G2 concentrations are presented in Figure 3.1. The product curves increase monotonically from zero, with all curves showing

Figure 3.1: Hydrolysis Time-Course Curves. pNP-G2 hydrolysis product (pNP) measured in time for 1.4 μM Cel5Acd (left) and 4.0 μM Cel5Acd (right). The initial substrate concentrations in each are: 1.0 mM (diamond), 1.5 mM (square), 2.0 mM (circle), and 2.5 mM (triangle). The solid curves represent fits to the data.


pNP (mM)
2.0
1.5
1.0
0.5
0.0
0
50
100
150
200
Time (min)


pNP (mM)
2.0
1.5
1.0
0.5
0.0
0
50
100
150
200
Time (min)

complete hydrolysis within six hours. In all cases, the final pNP product was less than the initial pNP-G2 substrate. For example, only about 78% of the substrate is converted to reporter for $S_T = 2.5$ mM and about 87% of the substrate is converted to reporter for $S_T = 1.0$ mM. This suggests that pNP-G2 had reacted by an alternate pathway where hydrolysis did not occur adjacent to the reporter group.

Products distribution. To identify the distribution of final products, 4 μM Cel5Acd and 5 mM pNP-G2 were reacted to exhaustion (4 hours), and the final solution was run on a TLC plate. The result, shown in Figure 3.2, confirms that the major products of pNP-G2 hydrolysis are pNP and G2, with minor products pNP-G1, G3, and trace amounts of G1. Figure 3.2 also reveals that no pNP-G2 was present in the final reaction, i.e. the substrate was reacted completely. This was validated by boiling the exhausted reaction under basic conditions (2M Na_2CO_3), which catalyzes the complete hydrolysis of any remaining pNP-G2 to G2 and pNP. This method (data not shown) confirmed that no pNP-G2 remained in the reaction at the final times in Figure 3.1.

Inhibition assays. To test for glucose or cellobiose inhibition, Cel5Acd /pNP-G2 reactions were loaded with increasing concentrations of G1 and G2 and allowed to react until an apparent final state was reached. The pNP yields for 1.7 μM Cel5Acd hydrolyzing 2 mM pNP-G2 after 4 hours with 0 mM to 3 mM G1 or G2 are presented in Figure 3.3. At these concentrations G2 does not influence the reaction yield or composition, whereas above certain concentrations, G1 shows strong inhibition. Compositions were determined with TLC (data not shown), revealing that a large fraction of pNP-G2 remained unreacted in the high-glucose assays.

Figure 3.2: Product Compositions. TLC of a 4 μM Cel5Acd and 5 mM pNP-G2 reaction after 4 hours of hydrolysis. By lane: (a) Glucose (G1) through cellopentaose (G5) standard, (b) the reaction, (c) pNP-G1, and (d) pNP-G2. Note that a faint G1 band in lane (b) is often observed but is barely visible on scanned images of the TLC plate.

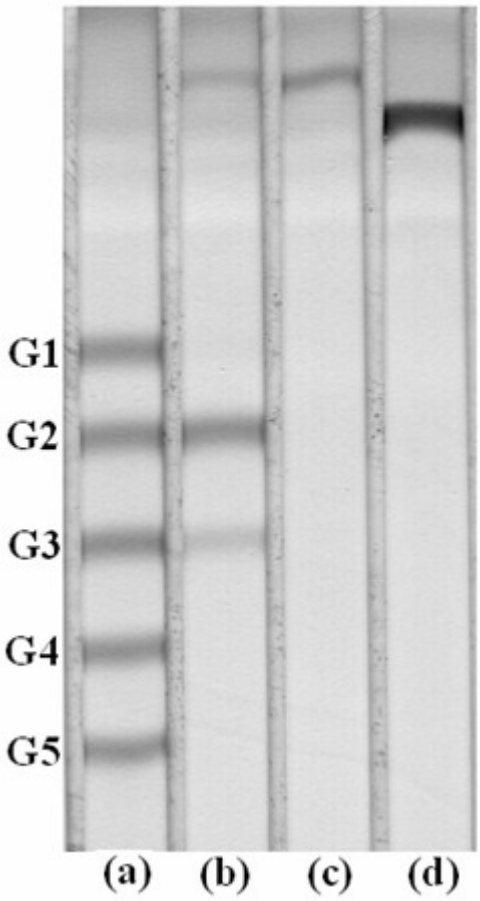
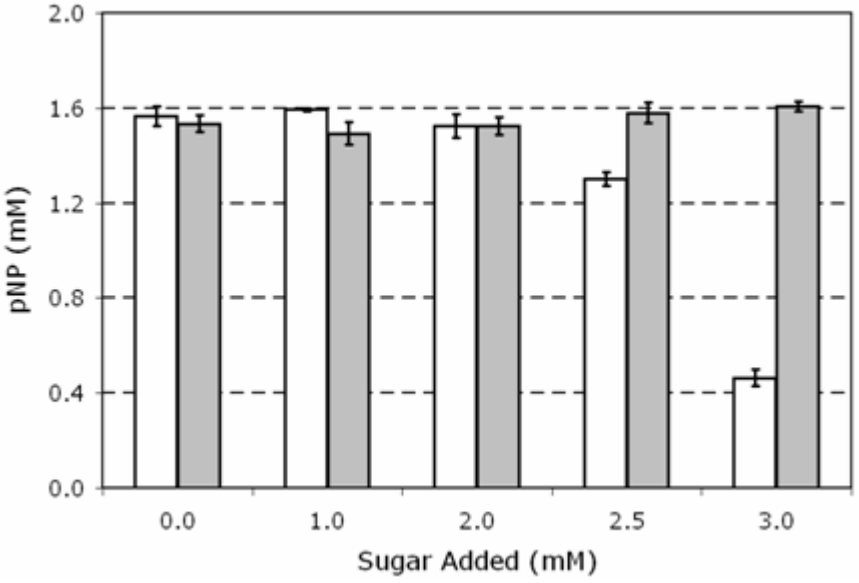

G1
G2
G3
G4
G5
(a)
(b)
(c)
(d)

Figure 3.3: Sugar Inhibition. Final concentrations of pNP product for reactions containing increasing concentrations of glucose (white) or cellobiose (grey). The initial concentrations of pNP-G2 substrate (2 mM) with Cel5Acd enzyme (1.7 μM) are the same in each. The reactions were conducted in triplicate and terminated after 4 hours.


2.0
1.6
1.2
0.8
0.4
0.0
pNP (mM)
0.0
1.0
2.0
2.5
3.0
Sugar Added (mM)

3.4 Discussion

3.4.1 Mechanism for Hydrolysis and Transglycosylation

Cel5Acd belongs to glycosyl hydrolase family 5-2, which share a common TIM barrel fold, hydrolyze β-glycosidic bonds via a double-displacement retaining mechanism, and contain eight invariant amino acids **[40, 142]**. The 3D molecular structure of Cel5Acd, solved using x-ray crystallography to 1.60 Å resolution for the wild-type and to 1.77 Å resolution for the inactive E355Q mutant with G4 bound, has been published in the Protein Data Bank (PDB: 2CKS, 2CKR). Comparison of this structure to those of other family 5-2s reveals that Cel5Acd's active site is a shallow cleft at the C-terminus of the α/β barrel with at least five pyranose ring-binding sites or "subsites" within the active site from -3 to +2 **[143-145]**. Subsites are numbered in the negative and positive directions toward the non-reducing and reducing end of the substrate, respectively, with hydrolysis always occurring between the +1 and -1 subsites **[146]**.

Next we develop a molecular-level mechanism for the hydrolysis and transglycosylation of pNP-G2 by Cel5Acd, involving the five observed subsites and an additional putative sub-site in the +3 location. The overall mechanism, shown in Figure 3.4, can be thought of as two pathways with a shared intermediate. The first pathway involves simple hydrolysis via the traditional double-displacement mechanism of glycosyl hydrolases **[42, 43, 131]**. Here, enzyme (*E*) and pNP-G2 substrate (*S*) combine reversibly to form a Michaelis complex (C_1), followed by a nucleophilic attack, which cleaves the aglyconic bond, releasing pNP reporter (*R*) and forming an α-glycosyl/enzyme intermediate (C_2). A water molecule then enters the active site and attacks the α-glycosyl/enzyme intermediate, cleaving the glycosyl-

Figure 3.4: Proposed Mechanism for Hydrolysis and Transglycosylation of pNP-G2 by Cel5Acd. The substrate and active site are presented in a simplified cartoon form and are not to scale. The locations of the -3 and +3 subsites are implied on the left and right side of the active site. Sites -3 through +2 are observed directly in the 2CKR crystal structure, whereas the existence of a +3 subsite is not yet proven. Glycosyl residues are shown as filled circles, and pNP residues are shown as empty circles.

E
Glu-355
Glu-263
-3
-2
-1
+1
+2
+3
(pNP-G2)
S
C1
(pNP)
R
C2
(G2)
P
C3
(G3)
P*
(pNP-G1)
R*

enzyme bond and releasing cellobiose product (*P*). In the second, transglycosylation pathway, the intermediate complex binds another substrate molecule and forms a second Michaelis complex (C_3). The C4 hydroxyl group at the non-reducing end of the second substrate acts as a nucleophile, breaking the glycosyl/enzyme bond and forming a non-covalently bound pNP-cellotetraose (pNP-G4) molecule. The pNP-G4 then shifts by one subsite in the non-reducing-end direction before being hydrolyzed via the double-displacement mechanism to secondary G3 product (*P**) and secondary pNP-G1 reporter (*R**). The actions of the catalytic nucleophile, Glu-355, and catalytic acid/base, Glu-263, are highlighted here, although the entire process includes many more active-site residues.

This mechanism assumes the initial cleavage event leading to the intermediate C_2 occurs strictly at the aglyconic bond. TLC data in Figure 3.2 confirms this specificity, since the product ratio of pNP-G1 (minor) to G1 (trace) is very large. Similar specificity has been documented for a homologous enzyme-substrate system **[147]**. The mechanism also assumes that pNP-G4 processively shifts in the non-reducing-end direction after formation. pNP-G4 could alternatively be released into solution, rebound, and hydrolyzed. pNP-G4 is not observed in solution by TLC, but its release would be difficult to detect because longer chain oligosaccharides are typically hydrolyzed much faster than shorter ones **[134, 148-151]**. Although either scenario is possible, only the processive one is depicted in Figure 3.4.

Cellobiose accumulates as a non-hydrolyzable product in solution and is known to be a strong competitive inhibitor of other cellulases **[152]**. Cellobiose could also act as an acceptor, competing with the pNP-G2 in the transglycosylation reaction, resulting in the formation of G4 instead of pNP-G4. The G4 would be difficult to detect if rapidly hydrolyzed to G2, but the more telling result would be an increase in pNP formation; G2 blocks the pNP-G2 transglycosylation pathway, ensuring more

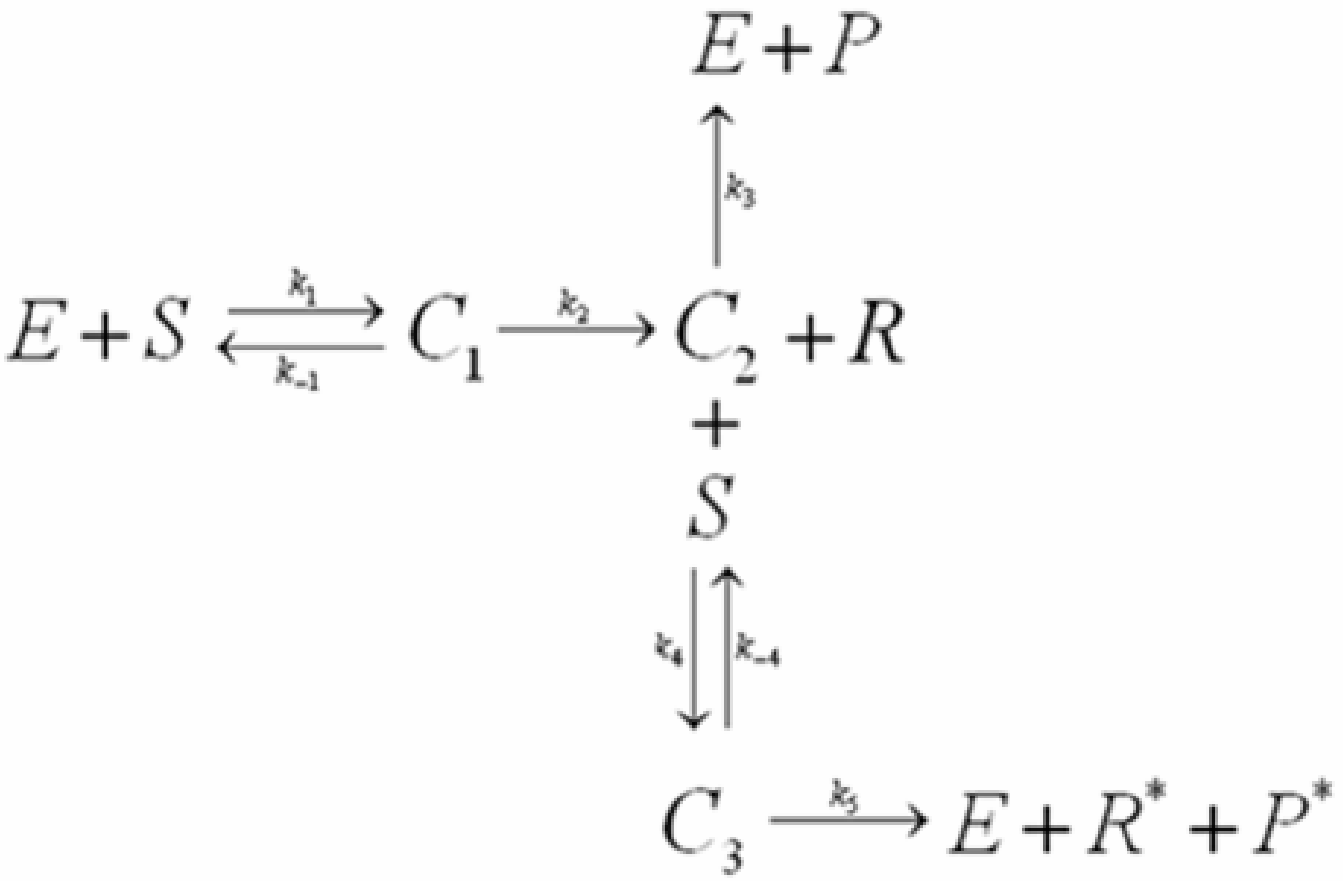
pNP-G2 is converted to pNP **[134]**. Interestingly, Figure 3.3 shows the pNP-yield is unaffected by G2 (up to 3mM), indicating that neither hydrolysis nor transglycosylation is inhibited here. Comparative assays with equimolar G1 (*cf*. Figure 3.3) revealed that glucose at high concentrations may be a stronger inhibitor than cellobiose for this enzyme, which is puzzling and perhaps worthy of its own investigation, because cellobiose is typically a stronger inhibitor than glucose. We do not include glucose inhibition in our mechanism, nor do we include the effects of the minor products G3 and pNP-G1, because all three species are present at very low concentrations in our measurements.

3.4.2 Mathematical Model for Hydrolysis Kinetics

In Figure 3.5 we present a network of elementary chemical reactions involving nine distinct chemical species, the minimum number necessary to account for the mechanism of Figure 3.4. For simplicity, we lump the entire transglycosylation event (both formation and hydrolysis of pNP-G4) into a single, non-reversible reaction. The reaction network of Figure 3.5 is the same as has been presented previously for several other enzyme-substrate systems that transglycosylate **[132, 133, 136]**.

Since there are nine species, nine mass balances (ordinary differential equations, ODEs) are required to describe the time-dependent concentrations of all the species in a closed batch reactor. If the reactor is charged initially with known concentrations of enzyme E_T and substrate S_T, there are four algebraic relationships that account for conservation of the total amounts of enzyme, pNP, and glycosyl molecular fragments in the reactor and for the stoichiometric necessity that $R^* = P^*$. Consequently, only five of the nine ODEs are independent. When substrate is present in large excess,

Figure 3.5: Reaction Network. Network of elementary chemical reactions corresponding to the mechanism of Figure 3.4, with k_i representing the various reaction rate constants. Each species is defined as follows: free Cel5Acd = *E*, pNP-G2 substrate = *S*, product G2 = *P*, secondary product G3 = *P**, reporter pNP = *R*, secondary reporter pNP-G1 = *R**, enzyme-substrate Michaelis complex = C_1, covalent α-glycosyl/enzyme intermediate =C_2, and the secondary Michaelis complex between substrate and α-glycosyl/enzyme intermediate = C_3.


E + P
k3
E + S
k1
k-1
C1
k2
C2 + R
+
S
k4
k-4
C3
k5
E + R* + P*

e.g., $\left(E_T \middle/ S_T \right) \cong 10^{-3}$ as in our experiments, some other simplifications are possible, and the five independent mass balances reduce to the following:

(1a) $$\frac{dC_1}{dt} = k_1\left(E_T - C_1 - C_2 - C_3\right)S - k_1 K_m C_1;$$

(1b) $$\frac{dC_2}{dt} = k_2 C_1 - k_3 C_2 - k_4 S C_2 + k_{-4} C_3;$$

(1c) $$\frac{dC_3}{dt} = k_4 S C_2 - k_4 K_m^* C_3;$$

(1d) $$\frac{dS}{dt} = -k_2 C_1 - k_5 C_3;$$

(1e) $$\frac{dR}{dt} = k_2 C_1.$$

The k_i's are rate constants for the various elementary reactions in Figure 3.5, and the parameters $K_m = \frac{k_{-1} + k_2}{k_1}$ and $K_m^* = \frac{k_{-4} + k_5}{k_4}$ are pseudo-Michaelis-Menten constants derived from them.

Equations (1a-e) are to be solved with the initial conditions $C_1(0) = C_2(0) = C_3(0) = R(0) = 0$ and $S(0) = S_T$, but they are nonlinearly coupled and cannot be solved in closed form. One might be tempted at this point to solve (1a-e) numerically while varying the seven rate constants k_i to optimize the agreement between the calculated $R(t)$ and the data of Figure 3.1. To do this is perilous, however, because the simple structure and statistical uncertainty (error bars) of the data in Figure 3.1 cannot possibly justify a curve-fit with seven degrees of freedom. It is likely one would find multiple parameter sets with poor fit-quality and very low statistical precision **[141]**.

We need a rational way, guided by reliable chemical insight, to simplify the equations and reduce the number of fitting parameters.

The kinetics problem can be simplified from a seven-parameter fit to a four-parameter fit and solved analytically by applying the quasi-steady state approximation (QSSA) to all three complex species, i.e. by setting $dC_i/dt = 0$ **[125, 126]**. Although the QSSA is assuredly valid for intermediate complexes in simple Michaelis-Menten reactions, it is not clearly valid for all three complexes in this case, since they are formed in sequence **[153]**. In the Appendix we derive inequality relationships that delimit when the QSSA applies, and we explain why it applies here. With the QSSA simplification for the complexes, equations (1a-e) reduce to the following four-parameter, coupled-pair of ODEs:

(2a) $$\frac{1}{E_T}\frac{dR}{dt} = \frac{\tilde{k}_{cat} S + \tilde{k}_{cat,2} S^2}{\tilde{K}_m + S + \frac{1}{\tilde{K}_{m,2}} S^2};$$

(2b) $$\frac{1}{E_T}\frac{dS}{dt} == \frac{\tilde{k}_{cat} S + 2\tilde{k}_{cat,2} S^2}{\tilde{K}_m + S + \frac{1}{\tilde{K}_{m,2}} S^2}.$$

The four lumped kinetics parameters in these are:

(3a) $$\tilde{k}_{cat} = \frac{k_3 K_m^*}{k_3 K_m^* + k_2 K_m^* + k_5 K_m} k_2;$$

(3b) $$\tilde{K}_m = \frac{k_3 K_m^*}{k_3 K_m^* + k_2 K_m^* + k_5 K_m} K_m;$$

(3c) $$\widetilde{K}_{m,2} = \frac{k_3 K_m^* + k_2 K_m^* + k_5 K_m}{k_2 + k_5};$$

(3d) $$\widetilde{k}_{cat,2} = \frac{k_2 k_5}{k_3 K_m^* + k_2 K_m^* + k_5 K_m} \quad .$$

The rate expression for reporter formation (2a) and the lumped parameters (3a-d) have been identified before for similar systems, and their combination has been called the "modified Michaelis-Menten equation" for reaction networks of this type **[132, 133, 136]**. Equations (2a-b) can be integrated via a procedure described in the Appendix to give $S(t)$ implicitly and $R(t)$ as a function of $S(t)$:

(4a) $$\widetilde{k}_{cat} E_T t = \frac{\omega}{\widetilde{K}_{m,2}}(S_T - S) - \widetilde{K}_m \ln\left(\frac{S}{S_T}\right) + \left[\widetilde{K}_m - \omega\left(1 - \frac{\omega}{\widetilde{K}_{m,2}}\right)\right] \ln\left(\frac{\omega + S}{\omega + S_T}\right);$$

(4b) $$R = \frac{1}{2}\left[(S_T - S) - \omega \ln\left(\frac{\omega + S}{\omega + S_T}\right)\right].$$

A new lumped parameter appears naturally in these as a consequence of the integration procedure:

(5) $$\omega = \frac{\widetilde{k}_{cat}}{2\widetilde{k}_{cat,2}} = \frac{k_3 K_m^*}{2k_5} = \frac{k_3}{2k_4}\left(1 + \frac{k_{-4}}{k_5}\right).$$

For complete hydrolysis, the pNP-group in the substrate becomes either free pNP reporter (*R*) or the aglyone residue in a non-hydrolyzable pNP-G1 secondary reporter (*R**). The final distribution of pNP in these two products, i.e. the selectivity

of the reaction for hydrolysis vs. transglycosylation, can be quantified by equation (4b) with $S \to 0$. One finds

$$(6) \qquad f = \frac{R_\infty}{S_T} = \frac{1}{2}\left[1 - \frac{\omega}{S_T}\ln\left(\frac{\omega}{\omega + S_T}\right)\right];$$

where f is the fraction of pNP in the initial substrate that is ultimately converted to *R*, and $1 - f$ is fraction of pNP in the substrate converted to R^*. When S_T is small compared to ω, $f \to 1$, all substrate appears as reporter, and the transglycosylation pathway is effectively extinguished. Conversely, when S_T is large compared to ω, $f \to \frac{1}{2}$, and *R* and *R** are produced in equal amounts. This is the maximum possible yield for the transglycosylation pathway, because one substrate molecule must act as a donor and hydrolyze in the self-transfer reaction before another can act as an acceptor and transglycosylate. We conclude that the single kinetics parameter ω, which has units of concentration and emerges only in the integral form of the rate expression, uniquely determines the intrinsic tendency of any substrate-enzyme system that follows the reaction network in Figure 3.5 to branch toward hydrolysis or transglycosylation products.

Equation (5) for ω provides quantitative guidance for mutating the enzyme's active site to change the product distribution. For example, mutations that increase k_3 relative to k_4 would directly favor hydrolysis over transglycosylation, whereas mutations that increase k_{-4} relative to k_5 would favor hydrolysis only insofar as the ratio k_{-4}/k_5 can be made large relative to one. Any such mutations at the active site would likely affect the magnitudes of k_1, k_{-1}, or k_2 also, but these only affect the overall rate of reaction, not the distribution of products.

When S_T is small compared to ω, measured $R(t)$ curves will evidence simple Michaelis-Menten kinetics. This can be seen by evaluating the asymptotic behavior of

equations (4a-b) for $S_T, S << \omega$. Expanding the logarithmic terms to leading order in S/ω and S_T/ω gives:

(7a) $$\bar{k}_{cat} E_T t = (S_T - S) - \bar{K}_m \ln\left(\frac{S}{S_T}\right);$$

(7b) $$R = S_T - S;$$

where

(8a) $$\bar{k}_{cat} = \frac{\tilde{k}_{cat}}{\left(1 - \frac{\tilde{K}_m}{\omega}\right)} = \frac{k_3 K_m^*}{\left(k_3 K_m^* + k_2 K_m^* - k_5 K_m\right)} k_2;$$

(8b) $$\bar{K}_m = \frac{\tilde{K}_m}{\left(1 - \frac{\tilde{K}_m}{\omega}\right)} = \frac{k_3 K_m^*}{\left(k_3 K_m^* + k_2 K_m^* - k_5 K_m\right)} K_m.$$

Equation (7a) duplicates exactly the functional form of the integrated Michaelis-Menten equation but with effective constants $\bar{k}_{cat}$ and $\bar{K}_m$ **[97]**. This reveals a general danger of investigating the hydrolysis activity of enzymes by performing experiments for a limited range of substrate compositions. If transglycosylation is in fact possible at high substrate concentrations, but one has the misfortune to accidentally select values of S_T that are small compared to ω for the system of interest, one might mistakenly conclude that only hydrolysis occurs, and that it occurs via a simple Michaelis-Menten mechanism with a single intermediate complex, in which case $\bar{k}_{cat}$ and $\bar{K}_m$ would be functions of only three elementary rate constants. In fact the mechanism may involve three intermediate complexes, in which case $\bar{k}_{cat}$ and $\bar{K}_m$ are

the more complicated functions of elementary rate constants given by equations (8a-b).

3.4.3 Data Analysis

Data were collected in a region of parameter space where $S_T \cong \omega$, i.e. where the transglycosylation yield varied considerably with S_T. All 104 data points for reporter concentration vs. time in Figure 3.1 (with their sample standard deviations) were fit simultaneously with equations (4a-b) using a non-linear, weighted least-squares approach, as described in the Experimental section. The best-fit values, standard deviations, and relative precisions for the fit-parameters $\widetilde{k}_{cat}$, $\widetilde{K}_m$, $\widetilde{K}_{m,2}$, and ω are presented in Table 3.1. We also calculate and present the dependent parameter $\widetilde{k}_{cat,2} = \widetilde{k}_{cat} / 2\omega$ for comparison, since this parameter appears naturally in the differential form of the mass balance for $R(t)$, equation (2a), used by others for initial-rate analysis in similar systems **[132, 133, 136]**. The solid lines in Figure 3.1 represent the kinetics-model solution with these parameters, which gives convincing agreement with the data for all values of E_T and S_T investigated here.

Equation (4b) reveals that the long-time amplitude of the $R(t)$ curves is a function of the single parameter ω. The copious long-time data in the right panel of Figure 3.1 ensure that ω can be specified to very high precision (±2.3%, *cf.* Table 3.1) by curve-fits to the data of Figure 3.1. Conversely, the parameters $\widetilde{k}_{cat}$, $\widetilde{k}_{cat,2}$, $\widetilde{K}_m$, and $\widetilde{K}_{m,2}$ cannot be determined with rewardingly high precision (±20-50%, *cf.* Table 3.1). These parameters only affect the shape of $R(t)$ where it rises from zero and bends toward its long-time, steady state value, and there are fewer data points in this regime. Extracting more parameters from fewer data ensures lower statistical precision.

Table 3.1. Best Fit Parameters. Best-fit values of parameters $\widetilde{k}_{cat}$, $\widetilde{K}_{m}$, $\widetilde{K}_{m,2}$, and ω, their standard deviations; and their relative precisions (= best-fit value ÷ standard deviation x 100%). Note that $\widetilde{k}_{cat,2}$ was calculated using equation (5). The solid lines in Figure 3.1 were generated using these values in equations (4a-b).

Parameter	Value	Rel. Precision
$\widetilde{k}_{cat}$ (min^{-1})	19.4 ± 4.6	24%
$\widetilde{k}_{cat,2}$ ($mM^{-1}min^{-1}$)	7.48 ± 1.8	24%
$\widetilde{K}_m$ (mM)	0.91 ± 0.3	33%
$\widetilde{K}_{m,2}$ (mM)	2.69 ± 1.4	52%
ω (mM)	1.30 ± 0.03	2.3%

The most important quantitative result of the integral curve-fits is the value of the selectivity-determining parameter $\omega = 1.30 \pm 0.03$ mM, since it can be used in equation (6) to predict the final product distribution as a function of initial substrate concentration. For example, reducing the substrate concentration to $S_T = 0.10$ mM for would give $f = 0.98$, i.e. almost pure hydrolysis products, and since $S_T/\omega = 0.10/1.30 \cong 0.08$ under these conditions, the pNP-vs.-time curves would evidence simple MM kinetics according to equation (7a-b). The apparent catalytic rate constant and MM constant would be $\overline{k}_{cat}$ = 65.5 min^{-1} and $\overline{K}_m$ = 3.08 mM according to equations (8a-b). Conversely, very high substrate concentrations are required to achieve a near-maximum yield of transglycosylation products for this system; for example, $S_T \cong 50$ mM is required to achieve $f \cong 0.55$.

3.5 Conclusions

Nitrophenyl glycosides are commonly used substrates for studying the reactivity of retaining glycosyl hydrolases, yet the deceptively simple kinetics of nitrophenol formation can be complicated by concurrent transglycosylation pathways. Herein we propose a two-pathway mechanism for the hydrolysis and transglycosylation of pNP-G2 by Cel5Acd, and from it we develop a rate expression for reporter formation. We provide quantitative justification for using the QSSA for three intermediate complexes when integrating the rate expression, and we fit the integral form of the rate to pNP-vs.-time data to extract values of several lumped kinetics parameters. One parameter, which emerges from the integration and is quantified very precisely by the curve-fits we present, uniquely determines the variation of hydrolysis vs. transglycosylation selectivity as a function of initial substrate concentration. We demonstrate that when the substrate composition is

sufficiently low, the two-pathway mechanism predicts simple Michaelis-Menten kinetics for hydrolysis, even though three intermediate complexes are involved in the reaction. Our kinetics model and mechanistic explanation should be useful for interpreting results for other retaining glycosyl hydrolase systems where hydrolysis and transglycosylation are evident.

3.6 Acknowledgements

This project was supported by the Initiative for Future Agriculture and Food Systems Grant no. 2001-52104-11484 from the USDA Cooperative State Research, Education, and Extension Service. The authors thank Dr. David Wilson, his staff, and his students at Cornell University for their assistance and advice throughout the experimental portion of this study.

3.7 Solution Appendix

The QSSA is a versatile method for finding approximate solutions to some kinetics problems **[125, 126]**. The QSSA assumes that in a specific region of parameter space and time, the rates of formation and destruction of a chemical species are effectively equal, such that the time-derivative of its concentration is nearly zero, which reduces an ODE into an algebraic equation. This condition is approximately met for intermediate complexes in many but certainly not all enzymatic biochemical reactions. To identify when such a simplification is actually valid, one must first re-scale the problem carefully to a dimensionless form where all variables are $O(1)$ [O(x) = "order of x" in an asymptotic sense] **[91]**. When this is accomplished, the dimensionless coefficient in front of each term quantifies its relative magnitude; the

QSSA is valid when the coefficient of a properly scaled time derivative is small compared to one **[153]**. Here we outline a procedure for re-scaling equation (1a-e) to delimit the validity of the QSSA for the complexes in the reaction mechanism we propose, and we use the QSSA to derive the mathematical model for hydrolysis kinetics we presented earlier as equations (4a-b).

First we scale the dependent variables such that their new, dimensionless representations are $O(1)$ for all time. The correct scale for R and S is clearly S_T. To determine the scaling factor β_i for each complex C_i, assume the enzyme achieves equilibrium between all three complex states before an appreciable amount of product is formed or substrate is consumed. Returning to equations (1a-c), let $C_i = \beta_i$, $\frac{dC_i}{dt} = 0$, and $S = S_T$, and then solve the coupled algebraic equations to find:

(A1a) $$\beta_1 = \frac{\mathrm{K} S_T}{(S_T + K_m)(S_T + \mathrm{K})} E_T ;$$

(A1b) $$\beta_2 = \frac{S_T K_m^*}{(S_T + K_m^*)(S_T + \mathrm{K})} E_T ;$$

(A1c) $$\beta_3 = \frac{S_T^2}{(S_T + K_m^*)(S_T + \mathrm{K})} E_T .$$

In these equations $\mathrm{K} = \left(\frac{S_T + K_m}{S_T + K_m^*} \right) \left(\frac{k_3}{k_2} K_m^* + \frac{k_5}{k_2} S_T \right)$. Introduce dimensionless representations of the dependent variables in (1) by substituting $X_i = \frac{C_i}{\beta_i}$, $Z = \frac{S}{S_T}$, and $Y = \frac{R}{S_T}$, and then rearrange the constants so all the terms on the right hand side of each equation are dimensionless and $O(1)$. The re-scaled representation of equation (1e) for reporter formation is of most interest to us, since $R(t)$ is the species we track in our measurements. That equation becomes

(A2) $$\left(\frac{S_T}{k_2\beta_1}\right)\frac{dY}{dt} = X_1.$$

The coefficient on the left-hand side has dimensions of time and represents the timescale for reporter formation. Define a scaled time variable $T = \frac{t}{t_R}$ with $t_R = \frac{S_T}{k_2\beta_1}$, and recast all five ODEs in terms of *T*. One finds:

(A3a) $$\varepsilon_1\left(\frac{dX_1}{dT}\right) = Z - \delta_1 X_1 - \gamma_1 X_1 Z - \gamma_2 X_2 Z - \gamma_3 X_3 Z;$$

(A3b) $$\varepsilon_2\left(\frac{dX_2}{dT}\right) = \delta_2 X_1 - \left(\frac{k_3 + k_4 S_T Z}{k_3 + k_4 S_T}\right) X_2 + \delta_3 X_3;$$

(A3c) $$\varepsilon_3\left(\frac{dX_3}{dT}\right) = X_2 Z - X_3;$$

(A3d) $$\frac{dZ}{dT} = -X_1 - \delta_4 X_3;$$

(A3e) $$\frac{dY}{dT} = X_1.$$

The new dimensionless coefficients on the right-hand side of these are $\gamma_i = \frac{\beta_i}{E_T}$ (i.e., the fraction of enzyme in each complex state), $\delta_1 = \frac{\beta_1 K_m}{S_T E_T}$, $\delta_2 = \frac{k_2\beta_1}{\beta_2(k_3 + k_4 S_T)}$, $\delta_3 = \frac{k_{-4}\beta_3}{\beta_2(k_3 + k_4 S_T)}$, and $\delta_4 = \frac{k_5\beta_3}{k_2\beta_1}$. One can easily verify by substituting equations (A1a-c) for β_i that all $\gamma_i, \delta_i \le 1$. The new dimensionless coefficients on the left-hand side, which are ratios of characteristic timescales for complex and reporter formation, are

(A4a) $$\varepsilon_1 = \frac{t_{C_1}}{t_R} = \left(\frac{E_T}{S_T + K_m}\right)\left(\frac{K}{K + K_D + S_T}\right)\left(\frac{\mathrm{K}}{S_T + \mathrm{K}}\right)^2;$$

(A4b) $$\varepsilon_2 = \frac{t_{C_2}}{t_R} = \left(\frac{k_3 K_m^* + k_5 S_T}{k_3 K_m^* + k_5 S_T + k_{-4} S_T} \right) \left(\frac{K_m^*}{S_T + K_m^*} \right) \left(\frac{E_T}{S_T + \mathrm{K}} \right);$$

(A4c) $$\varepsilon_3 = \frac{t_{C_3}}{t_R} = \left(\frac{k_2}{k_5 + k_{-4}} \right) \left(\frac{E_T}{S_T + K_m} \right) \left(\frac{\mathrm{K}}{S_T + \mathrm{K}} \right) .$$

Equations (A3a-e) are a scaled, dimensionless version of the governing equations (1a-e). The dimensionless concentration variables – X_1, X_2, X_3, Y, and Z – are $O(1)$ for all time; the dimensionless coefficients satisfy $\gamma_i, \delta_i \le 1$; and the dimensionless time variable $T = \frac{t}{t_R}$ is $O(1)$ for the timescale of our measurements; consequently, the parameters ε_1, ε_2, and ε_3 quantify the magnitudes of the derivative-terms on the left-hand side. By inspection of (A4a) and (A4b), $\varepsilon_1, \varepsilon_2 \le \frac{E_T}{S_T}$, and since $\frac{E_T}{S_T} \cong O\left(10^{-3}\right)$ in our experiments, the QSSA is assuredly valid for C_1 and C_2 in this analysis. Equation (A4c) reveals that the QSSA is valid for C_3 when $\left(\frac{k_2}{k_5 + k_{-4}} \right) \left(\frac{E_T}{S_T} \right) << 1$, or alternatively $\frac{k_2}{k_4 K_m^*} << 10^3$ for the conditions of our measurements. This ratio of rate constants is effectively a branching ratio for formation of reporter R vs. R^*. Since we find that R is the majority product in all of our measurements (*cf.* Figure 3.1), we are confident that this inequality constraint is met, hence the QSSA is also valid for C_3 in this analysis.

We invoke the QSSA for the complexes in the dimensionless equations by assuming $\varepsilon_1 = \varepsilon_2 = \varepsilon_3 = 0$, which converts equations (A3a-c) from ODEs to algebraic equations. Solving (A3a-c) for X_1 and X_3 as functions of Z and substituting into (A3d) and (A3e) gives the following:

(A5a) $$\frac{dZ}{dT} = -\frac{c_1 Z + 2c_2 Z^2}{c_3 + c_4 Z + c_5 Z^2};$$

(A5b) $$\frac{dY}{dT} = \frac{c_1 Z + c_2 Z^2}{c_3 + c_4 Z + c_5 Z^2}.$$

The scaled versions of the initial conditions for these equations are

$Z(T=0) = \frac{S(t=0)}{S_T} = 1$ and $Y(0) = \frac{R(0)}{S_T} = 0$, and the new constants are $c_1 = \frac{k_3 \beta_2}{k_2 \beta_1}$,

$c_2 = \frac{k_5 \beta_3}{k_2 \beta_1}$, $c_3 = \frac{k_3 K_m \beta_2}{k_2 S_T E_T}$, $c_4 = \frac{\beta_2}{k_2 E_T}\left(\frac{k_5 K_m \beta_3}{\beta_2 S_T} + k_2 + k_3\right)$, and $c_5 = \frac{k_5 \beta_3}{E_T}\left(\frac{1}{k_2} + \frac{1}{k_5}\right)$.

Equation (A5a) can be integrated to give $T(Z)$, and dividing (A5b) by (A5a) to eliminate dT gives an equation that can be integrated to find $Y(Z)$. Performing these operations and returning the results to dimensioned variables gives equations (4a) and (4b), presented in section 3.4.

CHAPTER 4

A PENETRATION MODEL FOR THE ENZYMATIC HYDROLYSIS OF DENSE CELLULOSE FIBERS

4.1 Introduction

Cellulose, the major component of plant biomass, can be enzymatically hydrolyzed by a class of modular enzymes known as cellulases, which break the cellulose down into soluble sugars that can be used as feedstock in the production of renewable commodities and fuels **[1]**. Reliable mechanistic models for the hydrolysis process, particularly in regard to the evolution of the solid substrate as it is degraded, are lacking but would prove valuable to the study of these enzyme-substrate systems.

Non-complexed cellulase systems **[75]** from both fungi and bacteria contain mixtures of cellulase enzymes with a broad range of activities and specificities that share a common two domain structure: a catalytic domain (CD) linked to a carbohydrate binding domain (CBM) via a flexible linker **[31]**. The distinct role of each domain highlights two fundamental steps in the hydrolysis process: The CBM binds to insoluble cellulose, fixing the entire enzyme at an insoluble solid-liquid interface, and the CD hydrolyzes glycosidic bonds within cellulose at that interface. Consequently, almost all kinetics models for the enzymatic hydrolysis of cellulose involve binding followed by hydrolysis. Additional mechanistic details regarding enzyme activity, such as product inhibition **[77-79]**, multi-enzyme activities and synergism **[80],** enzyme inactivation **[79],** non-productive or irreversible binding, and so-forth, are added later.

The complicated structure of insoluble cellulose particles makes it difficult to model changes in the substrate during hydrolysis. Even so-called "model" cellulose substrates such as Avicel and bacterial microcrystalline cellulose **[10]**, which are nearly pure cellulose and macroscopically uniform, present a microscopically heterogeneous, insoluble solid phase with varying surface structure. Most kinetics models ignore these microstructural complexities and assume the cellulase binding sites within cellulose are uniformly accessible and homogenously distributed throughout the liquid phase. By modeling the binding sites as homogenously distributed chemical species, the dimensionality of the problem is reduced from two phases to one; consequently, the kinetics of hydrolysis can be described with ordinary differential equations (ODEs) rather than partial differential equations (PDEs), which is a dramatic mathematical advantage. While Zhang **[75]** proposed the name "semi-mechanistic" for these models, they are more properly called "pseudo-homogenous" models.

Pseudo-homogenous models are mathematically expedient but are unable to explain how the bulk (or "mixing-cup") concentrations of bound enzyme and product that one typically measures for enzymatic hydrolysis of cellulose in a batch reactor are related to the evolving structure and density of the cellulose particles. To capture these structural effects, the binding sites must be distributed within insoluble solid particles with defined geometries rather than being treated as a homogeneous chemical species in solution. Unfortunately, not many such continuum or "distributed" binding-site models can be found in the literature, and those that can be found are overly simplistic. For example, Movagharnejad *et al*. described the hydrolysis of non-porous cellulose particles as occurring via etching inward from an outer surface **[154]**. Later they superimposed that solution on a particle containing a non-reactive fraction, which gives the impression of a shrinking core **[155, 156]**. Their approach is noteworthy

because it was the first to treat the hydrolysis of a uniform distribution of cellulose binding-sites in a geometrically simple shape and to account for changes in the shape as hydrolysis proceeds. Unfortunately, the model was derived from unrealistic assumptions about the interactions between cellulases and cellulose. For example, they assumed enzyme binding only occurs once at the onset of the reaction; enzymes are only removed by surface crowding; bound enzymes deactivate in an independent 1st order reaction; and only enzymes that are bound participate in a product-inhibition equilibrium.

Herein we present a distributed binding-site model based on more realistic assumptions about the cellulose-cellulase interaction, and we investigate how hydrolysis affects the shape and density of solid cellulose particles, in this case long, uniform cylindrical fibers. We show that a proper mathematical description of the heterogeneous, enzyme catalyzed reaction of cellulose particles in a well-mixed batch reactor is an unavoidably complicated affair, even for idealized cylindrical geometry and for restrictive, simplifying assumptions about the hydrolysis kinetics. Nonetheless, our analysis reveals some important and very fundamental principles about the relationship between mixing-cup concentrations and the evolving structure of solid cellulose particles that can be used to guide experiments and to avoid drawing unsubstantial conclusions from kinetics data.

4.2 Problem Statement

The analysis we present here is inspired by a recent investigation of enzymatic hydrolysis of nearly cylindrical and uniform-diameter cotton fibers, where data for cellulose conversion vs. time are matched to SEM and AFM measurements of the conversion-dependent morphological changes in the fibers **[157]**. The fibers were

approximately 100 μm long by 20 μm in diameter, in a dilute suspension (1% w/v), and hydrolyzed to ~90% reducing sugar after about 24 days of exposure. Analysis of the fiber microstructure during hydrolysis showed roughening and swelling at the liquid-solid interface, as indicated by cross-fracture formation between and within microfibrils. The authors also observed the formation and growth of pits on the fiber surface. These pits widened and grew to depths of 6 nm, 50 nm, and 300 nm after 6, 12, and 12-18 days of hydrolysis, respectively. The results show that hydrolysis begins at the outer surface of the fibers, opening the microstructure there and allowing cellulases to penetrate and carry the hydrolysis process toward the center of the fibers. Pit growth within the exposed regions of the fiber indicates that cellulase binding and hydrolysis occurs within a reactive zone in sub-layers of the fiber. Our aim in this analysis is to capture this phenomenon in a continuum mathematical model that can be used to predict and evaluate the effects of changing various parameters that control the overall hydrolysis rate, and to investigate how those changes can be interpreted, either by measuring the liquid phase concentrations of reactants and products or by monitoring the microstructural changes that occur during hydrolysis.

4.2.1 Proposed Model and Assumptions

We begin by assuming a two-phase enzyme-substrate reaction between a porous, solid, insoluble cellulose phase and perfectly mixed liquid phase containing cellulase enzymes and soluble sugar products. Moreover, we adopt the following assumptions regarding the solid phase:

- The cellulose substrate is a continuum of reactive binding sites to which individual cellulase enzymes can bind reversibly and with which the bound

enzymes can react, converting the binding sites to soluble sugars;

- The insoluble cellulose phase consists of a uniform distribution of long, smooth, cylindrical fibers, so long relative to their diameter that only changes in the radial direction need to be considered;
- The initial solid volume-fraction is small, so that hydrolysis causes negligible change in the total volume of the reaction mixture;
- The fibers are sufficiently dense to ensure the diffusivity of cellulase enzymes in the solid phase is initially zero.

Let σ and b represent unoccupied binding-sites and bound cellulases (occupied binding-sites), and let E and P represent the liquid-phase enzyme and product, respectively. There are three elementary steps in the reaction mechanism:

(1a) Binding: $E + \sigma \overset{k_1}{\rightarrow} b$;

(1b) Release: $b \overset{k_{-1}}{\rightarrow} E + \sigma$;

(1c) Hydrolysis: $\frac{1}{\alpha}\sigma(+b) \overset{k_2}{\rightarrow} P(+b)$.

The parameters k_1 (1/*time*), k_{-1} (*vol/mol-time*), and k_2 (1/*time*) are elementary rate constants for each reaction as shown. They describe the intrinsic, point-wise kinetics of the cellulase-cellulose interactions; consequently, their magnitudes depend on the activity of the cellulases and the molecular-scale structure of the cellulose, either of which could in principle be varied in carefully designed experiments. The $(+b)$-term on both sides of reaction (1c) depicts the action of a cellulase catalytic domain

hydrolyzing cellulose and releasing soluble sugars while the enzyme remains attached to the binding site. The stoichiometric coefficient $1/\alpha$ (*dimensionless*) accounts for the fact that binding sites contain multiple product molecules, so that α molecules of product must be removed to consume one binding site **[75]**.

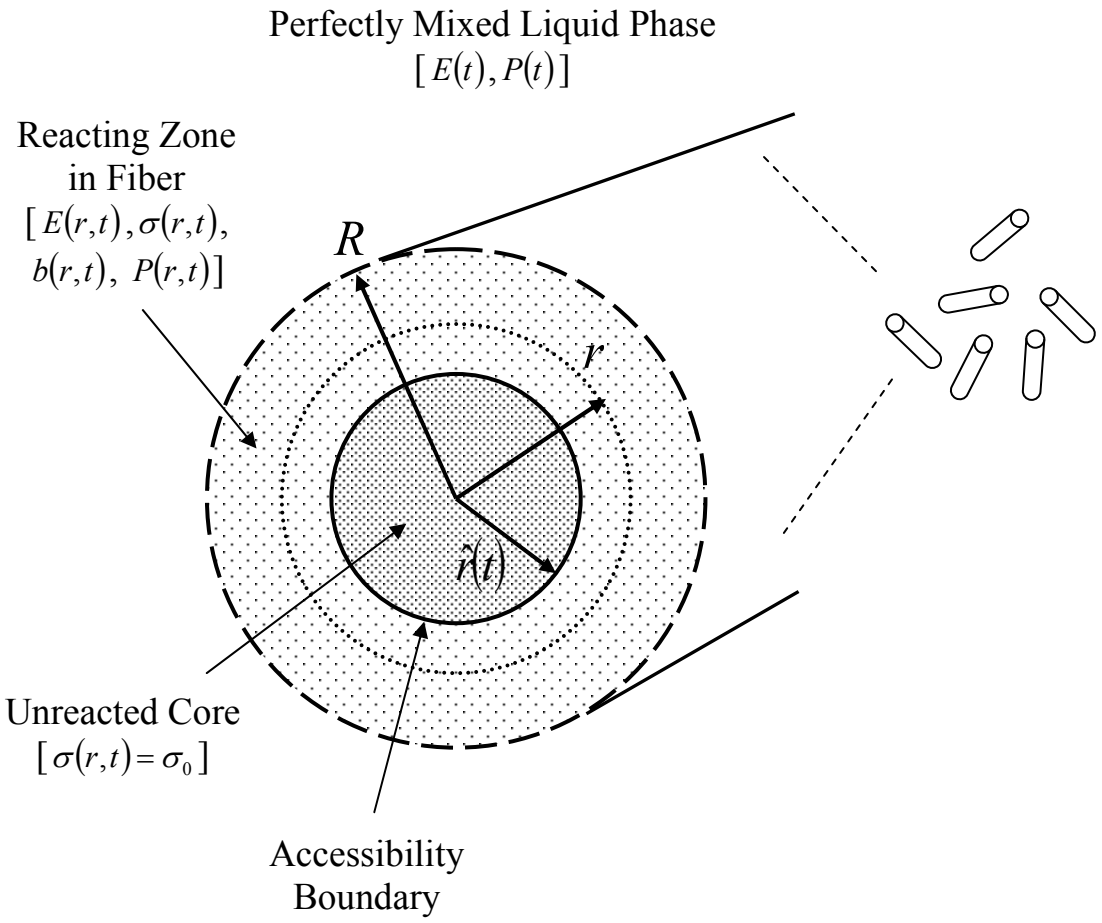
Consider a batch reactor charged with an initial enzyme concentration E_0 (*mol/vol*) and initial fiber concentration of M_0 (*mass/vol*), where each fiber has an initial radius R and an initial volumetric concentration of binding sites σ_0 (*mol/vol*). At $t = 0$ the enzymes bind at the outer surface ($r = R$) according to reaction (1a), and then release and hydrolyze according to reactions (1b) and (1c) (Figure 4.1). The hydrolysis reaction opens pits and voids, increasing the surface porosity until enzymes begin diffusing into the fiber, where they hydrolyze fresh cellulose, open more pores, and penetrate further into the fiber. A reacting zone, shaped like a cylindrical shell, forms between the outer radius of the fiber and an inner accessibility boundary where the cellulose porosity remains too small for the enzymes to penetrate, and over time this zone expands toward the center of the fiber. Let $r = \hat{r}(t)$ define the location of the boundary between the reacting zone and the as-yet unreacted substrate.

The dynamic evolution of the reacting zone ensures that localized values of porosity, surface area, species concentration, mobile species diffusivity, and the point-wise reaction rates will all vary as a function of time and radial position in the fiber. The mass balances for free enzyme, bound enzyme, and cellulose in the reacting zone can be described with the following PDEs:

(2a) $$\frac{\partial}{\partial t}E(r,t) = \frac{1}{r}\frac{\partial}{\partial r}\left[rD(r,t)\frac{\partial}{\partial r}E(r,t)\right] - R_B(r,t) + R_R(r,t);$$

(2b) $$\frac{\partial}{\partial t}b(r,t) = R_B(r,t) - R_R(r,t);$$

Figure 4.1: Fiber Hydrolysis Model. Dense, slender, cylindrical fibers in a dilute suspension are exposed to hydrolytic enzymes. An accessibility boundary located at $r = \hat{r}$ moves from the outer edge of the fiber toward to the center, behind which binding sites are exposed to enzymes and hydrolysis proceeds.


Perfectly Mixed Liquid Phase
$[E(t), P(t)]$
Reacting Zone
in Fiber
$[E(r,t), \sigma(r,t),$
$b(r,t),\ P(r,t)]$
R
r
$\hat{r}(t)$
Unreacted Core
$[\sigma(r,t) = \sigma_0]$
Accessibility
Boundary

(2c) $$\frac{\partial}{\partial t}\overline{\sigma}(r,t) = -R_H(r,t).$$

Notice that we have introduced an auxiliary point-wise concentration variable $\overline{\sigma}(r,t) = \sigma(r,t) + b(r,t)$ (*mol/vol*), the total concentration of binding sites either unoccupied or occupied, along with three instantaneous point-wise reaction rates (*mol/vol-time*): for binding, $R_B(r,t) = k_1 E(r,t) A(r,t)$; for release, $R_R(r,t) = k_{-1} b(r,t)$; and for hydrolysis, $R_H(r,t) = \frac{k_2}{\alpha} b(r,t)$. The release and hydrolysis reactions are unimolecular reactions that involve only bound reactants, hence their rate functions have simple mass-action forms. The binding reaction is bimolecular and depends on $A(r,t)$, the accessible area for enzyme binding per unit volume in the reacting zone. The parameter D in (2a) is the effective diffusivity for fluid-phase transport of enzyme within the reacting zone.

To solve the coupled mass balances (2a-c) we require a reliable model for the effect of reaction-extent on enzyme diffusivity and on the microstructure of the porous solid material, i.e. how $A(r,t)$ and $D(r,t)$ vary as a function of $\sigma(r,t)$. Problems of this general type, where a porous solid material is consumed and has its microstructure altered by a reaction that occurs at its interface with a fluid phase, have been analyzed before in the reaction engineering literature. Szekeley and Evans presented two microstructural models of reacting porous solids: one being an array of parallel cylindrical pores that grow in diameter as reaction proceeds, the other an array of touching spheres that shrink in radius but remain in position as reaction proceeds, and they solved coupled diffusion-reaction models for these materials numerically **[158]**. Gavalas developed a more sophisticated and realistic model that involves randomly oriented, intersecting cylindrical pores that grow in diameter and increase in interconnectedness as reaction proceeds, and he derived an analytical method for

calculating the speed of the inwardly penetrating diffusion/reaction front in spherical **[159]** pellets if the pore area and diffusivity are known functions of the reaction-extent. More relevant to the discussion here, Brem and Brouwers extended Gavalas's analysis by developing asymptotic solutions to the coupled diffusion/reaction equations that show how the front-speed behaves in certain limiting circumstances **[160]**.

Unfortunately, the effect of partial-hydrolysis on the microstructure of typical cellulosic materials has not been investigated to the extent necessary to guide us in developing a reliable microstructural model for this problem. However, a comparison of characteristic time scales for diffusion and reaction reveals that our problem is a peculiar limiting case for which an important simplification can be exploited to allow an analytical solution of the mass balances (eqs. 2a-c).

Consider first the characteristic timescale for the hydrolysis reaction (eq. 1c). Enzymatic hydrolysis measurements for bacterial microcrystalline cellulose, likely the most open, highest-surface-area form of pure cellulose available, show that times well in excess of 10 hrs are typically required to achieve >50% conversion of the cellulose to soluble sugar. Similar conversion times are found for other insoluble cellulose substrates. Therefore, a reasonable estimate for the characteristic timescale for hydrolysis of accessible cellulose is $\theta_{\bar{\sigma}} \geq 10$ hrs.

Now consider the diffusion process. A simple scaling analysis of Fick's law reveals that the characteristic time for diffusion $\theta_D = \frac{L^2}{D}$, where L is the diffusion path-length. In the situation we are addressing here, the maximum path length for diffusion is the radius of the fiber, which is $R \cong 10$ μm for naturally occurring fibers like cotton **[157]**. For this path-length, an enzyme diffusivity of $D \cong 3x10^{-11}$ cm^2/s would be required for the diffusion and hydrolysis timescales to be comparable, i.e. for $\theta_D = \theta_\sigma \cong 10$ hrs. If the effective diffusivity for cellulases in cellulose is

significantly less than this value, the rate of the overall hydrolysis process can be diffusion-limited, and a strong gradient in the fluid-phase enzyme concentration can exist in the reacting zone, at least in the worst-case scenario where the boundary is near the center and the diffusion path-length is as long as possible. If the effective diffusivity is significantly greater than this value, the overall process will be limited by hydrolysis kinetics alone, and an enzyme concentration gradient cannot be sustained in the reacting zone.

Now what can be said about the effective diffusivity of cellulases in cellulose? Theoretical treatments of hindered diffusion of rigid spheres in tightly constrained, fluid-filled pores show that the diffusivity approaches zero as the sphere radius approaches the pore radius, just as one would expect, but the diffusivity is enhanced significantly if the molecule is flexible **[161]**. Measurements for variable-length flexible polymers in porous glass show that the effective diffusivity in the pores remains greater than roughly 10^{-2} of the Stokes (free-fluid) diffusivity even when the hydrodynamic radius of the polymer molecule is comparable to the pore diameter **[161, 162]**. The Stokes diffusivity of cellulase enzymes is on the order of 10^{-7} cm^2/s **[163]**, and they are relatively flexible, so one expects the diffusivity of cellulases in tightly constrained pores of cellulose to be on the order of 10^{-9} cm^2/s or larger. (Careful measurements of D for partially hydrolyzed cellulose materials would be very useful in this context.) The corresponding diffusion time for fibers with $R \cong 10$ μm is $\theta_D \leq 0.3$ hrs. Since this is significantly smaller than the characteristic reaction time for cellulose $\theta_\sigma \geq 10$ hrs, we surmise that there will not be a strong gradient in the fluid-phase enzyme concentration in the reacting zone of the fiber.

Alternatively, one can express this same argument in terms of the classical Thiele modulus ϕ, which quantifies the relative importance of diffusion and kinetics limitations for reactions in porous media. A scaling analysis of the governing

equations, to be presented later, shows that the characteristic hydrolysis time is described by $\theta_{\bar{\sigma}} = \frac{\alpha\sigma_0}{k_2\beta}$, where $\frac{k_2}{\alpha}$ is the effective unimolecular rate constant for the hydrolysis reaction (eq. 1c, above), β is the concentration of bound enzyme achieved when the binding and release reactions (eqs. 1a-b) approach equilibrium, and σ_0 is the concentration of binding sites in native, unreacted cellulose. The Thiele modulus for this problem is then **[160],**

$$(3) \qquad \phi = \sqrt{\frac{\theta_D}{\theta_\sigma}} = R\sqrt{\frac{k_2\beta}{D\alpha\sigma_0}}.$$

Note that ϕ varies throughout the material since D increases as the hydrolysis reaction proceeds. In regions where $\phi >> 1$ there is a strong concentration gradient for the diffusing reactant, and diffusion-resistance limits the overall rate. The concentration gradient relaxes and kinetics-resistance limits the overall rate where $\phi << 1$.

Within the unreacted core $\phi \to \infty$ because $D = 0$ there. Brem and Brouwers have proven that this limiting circumstance ensures that the reaction boundary will penetrate at a constant speed **[164]**. In effect, the boundary is a shrinking core within the fiber, since hydrolysis must occur for the boundary to advance. When the penetrating boundary arrives at any point via the onset of hydrolysis, even a miniscule increase in the diffusivity from zero, for example to $D = 10^{-9}$ cm^2/s, decreases the Thiele modulus dramatically to $\phi \cong 0.2$. (Small changes in D have an unusually large impact on ϕ in this circumstance because R is so small and $\theta_{\bar{\sigma}}$ is so large.) The consequence is the fluid-phase enzymes can "chase" the penetrating boundary via diffusion with only a small concentration gradient, or alternatively, a significant concentration gradient can only be sustained in a vanishingly thin layer immediately behind the boundary.

In summary, one has the following scenario: there is a sharp boundary separating as-yet-unreacted from reacting material; the boundary moves at constant speed from the outer surface toward the center of the fiber; and the fluid-phase enzyme concentration is nearly constant behind the boundary. The latter circumstance allows us to consider the enzyme concentration as a function of time only, i.e. $E(r,t) \to E(t)$, so eq. (2a) drops out of the problem. Similarly, $P(r,t) \to P(t)$. Furthermore, we assume the accessible surface area for binding is proportional to the local concentration of unoccupied binding sites, so the mass-action form of the binding rate becomes $R_B = k_1 E \sigma$. We will show below that the mass balances admit a revealing analytical solution in this limiting circumstance. However, since we lack a detailed model for the microstructure of the cellulose that would allow us to calculate the accessible area for enzyme binding and the pore diffusivity as functions to the reaction extent, we cannot calculate the boundary speed from first principles. Instead, we will treat it as a variable parameter and investigate its effects on the nature of the solution. Since the boundary speed is a function of the microstructure of the cellulose, it could in principle be varied independently in experiments by choosing different cellulose sources or pretreatment strategies.

Let v be the boundary speed, so that $\hat{r} = R - vt$ defines the instantaneous location of the boundary. Thus, penetration is complete at $t = \theta_R = v/R$, and the fiber is completely exposed to enzyme and $\hat{r} = 0$ when $t > \theta_R$. Let $t' = t - (R-r)/v$ be the shifted or "local" reaction time, which is zeroed at the moment the accessibility boundary passes each radial position in the fiber (i.e. when the localized reaction begins.) Incorporating all of our assumptions up to this point, the reacting zone mass balance equations (2b-c) reduce to:

(4a) $$\frac{\partial b}{\partial t'} = k_1 E(\overline{\sigma} - b) - k_{-1} b ;$$

(4b) $$\frac{\partial \overline{\sigma}}{\partial t'} = -\frac{k_2}{\alpha} b \, .$$

The above PDE set must be solved at each position in the reacting zone with the initial conditions $b(r, t' = 0) = 0$ and $\overline{\sigma}(r, t' = 0) = \sigma_0$. The remaining mass balances require integration over the spatial dimensions of the reacting zone. Conservation of enzyme and a mass balance for the hydrolysis reaction yield the following expressions for the mixing-cup concentrations (*mol/vol*) of free enzyme, bound enzyme, and soluble sugar product, respectively:

(5a) $$E = E_0 - B \, ;$$

(5b) $$B = 2\psi \frac{1}{R^2} \int_{\hat{r}}^{R} br dr \, ;$$

(5c) $$\frac{1}{\alpha} P = 2\psi \frac{1}{R^2} \int_{\hat{r}}^{R} (\sigma_0 - \overline{\sigma}) r dr \, .$$

The parameter $\psi = \frac{M_0}{\rho_c}$ (*dimensionless*) with ρ_c (*mass/vol*) as the bulk-density of cellulose is the initial volume fraction of solid cellulose in the reactor.

The combination of equations (4a-b) for the concentration profiles in the reacting zone, their initial conditions and equations (5a-b) for the mixing-cup concentrations constitutes a complete statement of the problem. These equations are nonlinear, coupled, and complicated by the presence of the moving accessibility boundary, which makes even numerical solution difficult. We demonstrate in the next section that an accurate analytical solution can be obtained under certain relevant circumstances.

4.2.2 Convert to Dimensionless Representation

Define the dimensionless variables $x(r,t)=\frac{b(r,t)}{\beta}$, $\bar{z}(r,t)=\frac{\bar{\sigma}(r,t)}{\sigma_0}$, $W(t)=\frac{E(t)}{E_0}$, $X(t)=\frac{B(t)}{\psi\beta}$, and $Y(t)=\frac{P(t)}{\alpha\psi\sigma_0}$. From equation (4a), $\beta=\frac{E_0\sigma_0}{E_0+K_D}$ (*mol/vol*) is the leading-order estimate of maximum $b(r,t)$, where $K_D=\frac{k_{-1}}{k_1}$ (*mol/vol*). The characteristic timescale for enzyme binding is $\theta_b=\frac{\beta}{k_1E_0\sigma_0}$ and the characteristic timescale for total binding site reduction (hydrolysis) is $\theta_{\bar{\sigma}}=\frac{\alpha\sigma_0}{k_2\beta}$. Let $T=\frac{t}{\theta_{\bar{\sigma}}}$, $T'=\frac{t'}{\theta_{\bar{\sigma}}}$, $\rho=\frac{r}{R}$, and $\hat{\rho}=\frac{\hat{r}}{R}$. Equations (4a-b) become:

(6a) $\varepsilon\frac{\partial x}{\partial T'}=(1-\xi X)(\bar{z}-\mu x)-(1-\mu)x$, with $x(\rho,T'=0)$;

(6b) $\frac{\partial\bar{z}}{\partial T'}=-x$, with $\bar{z}(\rho,T'=0)=1$.

The dimensionless parameters $\mu=\frac{\beta}{\sigma_0}$ and $\xi=\frac{\psi\beta}{E_0}$ define the maximum fractions of binding sites occupied and of enzyme bound to substrate, respectively. The dimensionless parameters $\varepsilon=\frac{\theta_b}{\theta_{\bar{\sigma}}}$ and $\delta=\frac{\theta_R}{\theta_{\bar{\sigma}}}$ are timescale ratios for binding-to-hydrolysis and penetration-to-hydrolysis, respectively. Equations (5a-c) become:

(7a) $W = 1 - \xi X$;

(7b) $X = 2\int_{\hat{\rho}}^{1} x\rho d\rho$;

(7c) $Y = 2\int_{\hat{\rho}}^{1} (1 - \bar{z})\rho d\rho$.

For $T < \delta$, the accessibility boundary is located at $\hat{\rho} = 1 - \frac{1}{\delta} T$, where $T' = T - \delta(1 - \rho)$. For $T \geq \delta$, the fiber is completely penetrated, and $\hat{\rho} = 0$.

Equations (6a-b) fully characterize the system. Once solutions for $x(\rho, T')$ and $\bar{z}(\rho, T')$ are found, all remaining variables of interest can be generated from the integrals (7b-c) or from algebraic relationships.

4.3 Perturbation Solution for $0 < \xi << 1$

The governing equations (6a-b) are coupled, nonlinear, partial integro-differential equations that do not admit a closed-form solution. The scaling ensures the dependent variables are *O*(1) for all time, so that the relative size and importance of each term can be determined by relative size of the various dimensionless groups. Most notably, in the limit of $\xi \to 0$, corresponding to nearly constant free-enzyme concentration in the liquid phase during uptake and hydrolysis, equations (6a-b) become linear and soluble. The $\xi = 0$ solution can be used to generate a less restrictive analytical solution via a regular perturbation in ξ . The perturbation solution accounts approximately for partial depletion of the liquid-phase enzyme concentration caused by uptake in the cellulose.

The method we use here is a standard tool of engineering analysis and is described in many advanced math textbooks **[91, 102]**. One begins by postulating series expansions for the relevant dependent variables as follows,

(8a) $$x(\rho,T') \approx x_0(\rho,T') + \xi x_1(\rho,T') + O(\xi^2);$$

(8b) $$\bar{z}(\rho,T') \approx \bar{z}_0(\rho,T') + \xi \bar{z}_1(\rho,T') + O(\xi^2).$$

The corresponding mixing-cup concentrations of these variables are found by integrating over position, resulting in the following series expansions,

(9a) $$X(T) \approx 2\int_{\hat{\rho}}^{1} x_0 \rho d\rho + \xi 2\int_{\hat{\rho}}^{1} x_1 \rho d\rho + O(\xi^2) = X_0(T) + \xi X_1(T) + O(\xi^2);$$

(9b) $$\bar{Z}(T) \approx 2\int_{\hat{\rho}}^{1} \bar{z}_0 \rho d\rho + \xi 2\int_{\hat{\rho}}^{1} \bar{z}_1 \rho d\rho + O(\xi^2) = \bar{Z}_0(T) + \xi \bar{Z}_1(T) + O(\xi^2).$$

Inserting the expansions into (6a-b) and collecting terms by order in ξ generates the following sequence of equations at $O(1)$:

(10a) $$\varepsilon \frac{dx_0}{dT'} = \bar{z}_0 - x_0, \text{ with } x_0(T'=0) = 0;$$

(10b) $$\frac{d\bar{z}_0}{dT'} = -x_0, \text{ with } \bar{z}_0(T'=0) = 1;$$

and at $O(\xi)$:

(11a) $\varepsilon \frac{dx_1}{dT'} = \bar{z}_1 - x_1 - (\bar{z}_0 - \mu x_0) X_0$, with $x_1(T'=0) = 0$;

(11b) $\frac{\partial \bar{z}_1}{\partial T'} = -x_1$, with $\bar{z}_1(T'=0) = 0$.

Though tedious, solving these PDE sets is straightforward, as they are linear and coupled only within their respective orders in ξ. Assuming $\varepsilon < \frac{1}{4}$, the solution at $O(1)$ is,

(12a) $\bar{z}_0 = \frac{1}{\omega}\left(\frac{1}{\lambda_2} e^{\lambda_2 T'} - \frac{1}{\lambda_1} e^{\lambda_1 T'}\right)$;

(12b) $x_0 = \frac{1}{\omega}\left(e^{\lambda_1 T'} - e^{\lambda_2 T'}\right)$.

The "slow" eigenvalue at $O(1)$, which quantifies the long timescale for hydrolysis, is $\lambda_1 = -\frac{1}{2\varepsilon}(1-\omega)$, and the "fast" eigenvalue, which quantifies the short timescale for binding, is $\lambda_2 = -\frac{1}{2\varepsilon}(1+\omega)$, where $\omega = \sqrt{1-4\varepsilon}$. The $O(\xi)$ solution to equations (11a-b) and the mixing-cup concentrations at $O(\xi)$ are provided in Appendix 4.8.

4.4 Results

The point-wise concentrations of total binding sites and bound enzyme within the reacting zone, given by equations (8a-b), depend on four dimensionless model parameters (μ, δ, ε, and ξ) that can be varied independently in the model. However, the overall structures of the radial concentration profiles depend predominantly and most revealingly on the magnitudes of the timescale ratios δ and ε. For the purpose of

this discussion we assume $\varepsilon << 1$, based on the experimental observations that cellulase enzymes tend to bind their substrate much faster than they hydrolyze it. Consequently, the magnitude of the remaining timescale ratio, δ, uniquely governs the structure of the radial concentration profiles. Three special case reaction-penetration scenarios can be identified (Table 4.1).

If the timescale for penetration is short compared to the timescale for hydrolysis, for example $\delta = 0.1$, the homogeneous limit is approached, where complete penetration occurs before hydrolysis, and the concentration profiles appear uniform across the radius of the fiber (Figure 4.2, Panel A). If the timescales for penetration and hydrolysis are comparable, enzyme hydrolyzes substrate and penetrates the fiber concurrently (Figure 4.2, Panel B for $\delta = 1$). If hydrolysis is rapid compared to penetration, for example $\delta = 10$, the shrinking-core limit is approached, where extensive hydrolysis occurs before the accessibility boundary moves forward, resulting in a rapid drop in the concentration profile behind the boundary (Figure 4.2, Panel C). Clearly, these radial concentration profiles are unambiguous indicators of the reaction mechanism, as they reveal whether the penetration process or the point-wise kinetics of hydrolysis are controlling the overall rate of consumption of the fiber.

Unfortunately, this is not the case for the mixing-cup concentrations, the quantities typically measured in hydrolysis experiments: the curves for the mixing-cup concentrations of bound enzyme and product have similar overall shapes for all three penetration scenarios in Table 4.1. In every case the product curve increases monotonically from zero to one (Figure 4.3, Panel B), and the bound enzyme curve rises to an initial maximum at a time that is short compared to the total time over which the fiber is consumed, and then gradually decays toward zero over the same

Table 4.1: Reaction-Penetration Scenarios. Three reaction-penetration scenarios for fiber hydrolysis as determined by the relative sizes of the three characteristic timescales.

Name	**Relative Timescales**	**Timescale Ratios**
Homogenous	$\theta_R << \theta_b << \theta_{\bar{\sigma}}$	$\delta << \varepsilon << 1$
Concurrent	$\theta_b << \theta_R \sim \theta_{\bar{\sigma}}$	$\varepsilon << \delta \sim 1$
Shrinking Core	$\theta_b << \theta_{\bar{\sigma}} << \theta_R$	$\varepsilon << 1 << \delta$

Figure 4.2: Radial Concentration Profiles. Radial concentration profile for $\bar{z}(\rho,T)$ (left) and $x(\rho,T)$ (right) for three different values of the penetration-to-hydrolysis timescale ratio: Panel A: $\delta = 0.1$; Panel B: $\delta = 1$; Panel C: $\delta = 10$. The respective values of time, *T*, are reported adjacent to each distribution curve. In all cases $\mu = 0.5$, $\varepsilon = 0.02$, and $\xi = 0.1$.

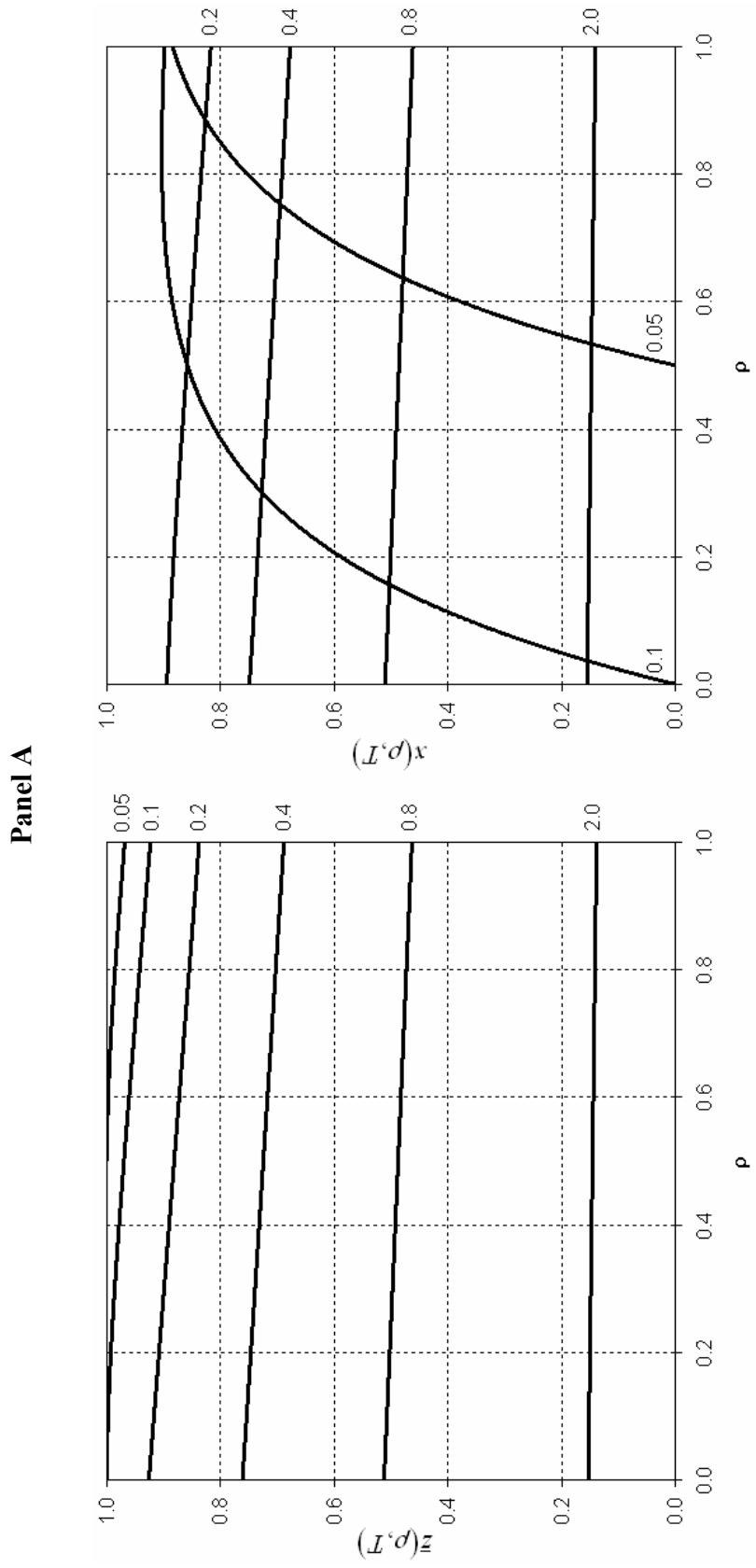

Panel A
$x(\rho,T)$
$\pi(\rho,T)$
ρ
0.05
0.1
0.2
0.4
0.8
2.0
0.0
0.2
0.4
0.6
0.8
1.0

Figure 4.2 Continued

Panel B

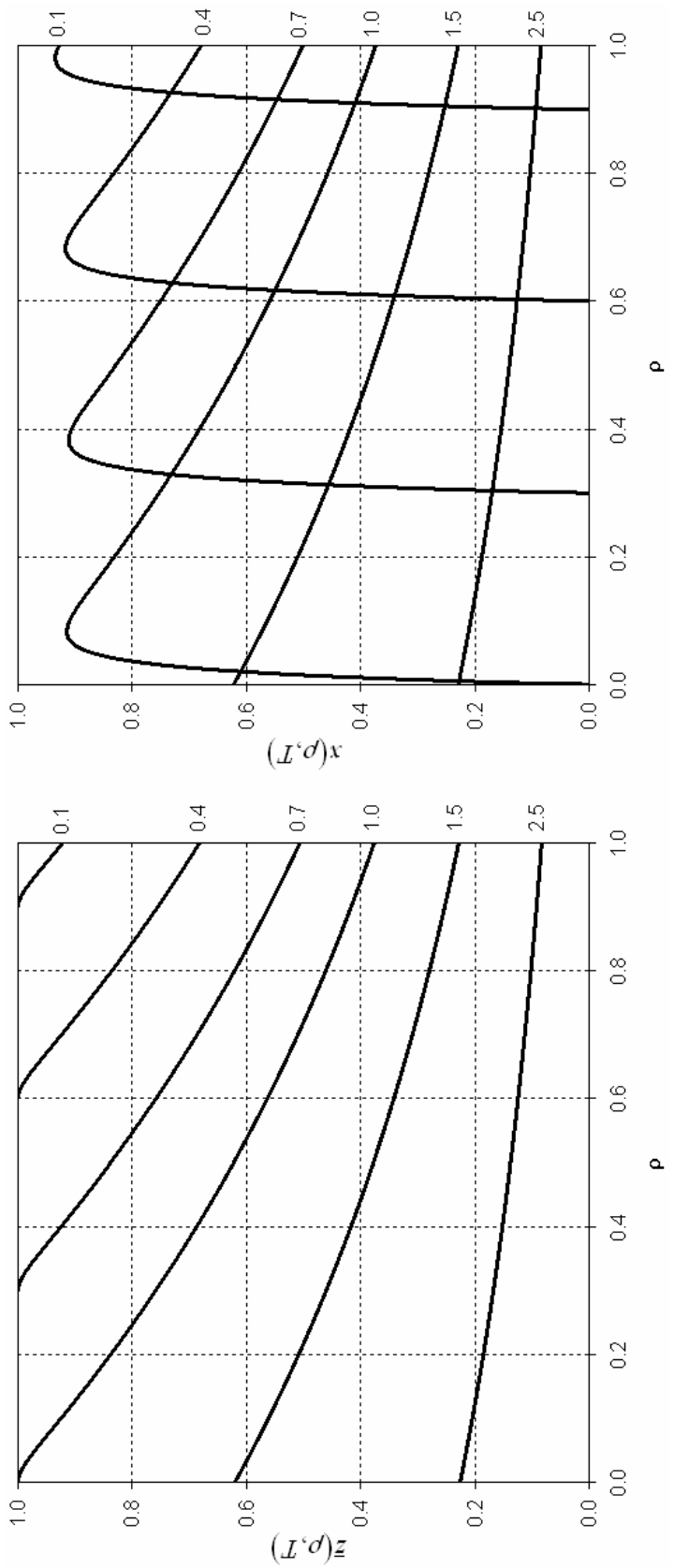


Figure 4.2 Continued

Panel C

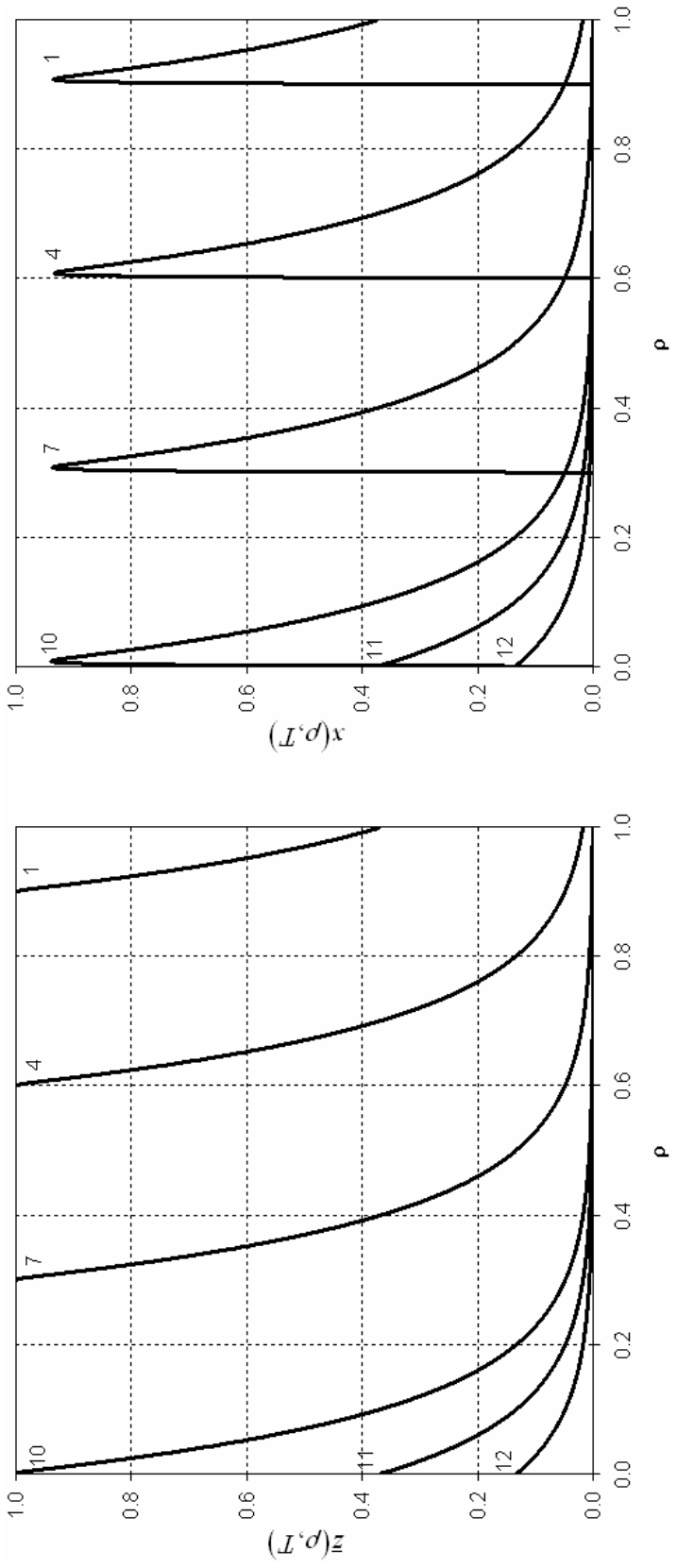


Figure 4.3: Bound Enzyme & Product Concentrations. Mixing-cup bound enzyme (Panel A, $B(t)/B_{\max}$) and product (Panel B, $P(t)/P_{\max}$) concentrations for three different values of the penetration to hydrolysis timescale ratio, δ: $\delta = 0.01$ (dotted), $\delta = 1$ (solid), and $\delta = 100$ (dashed). The values of time (x-axis) are scaled relative to $t_{1/2}$, the value of time required to convert half of the substrate to product. In all cases $\mu = 0.5$, $\varepsilon = 0.02$, and $\xi = 0.1$.

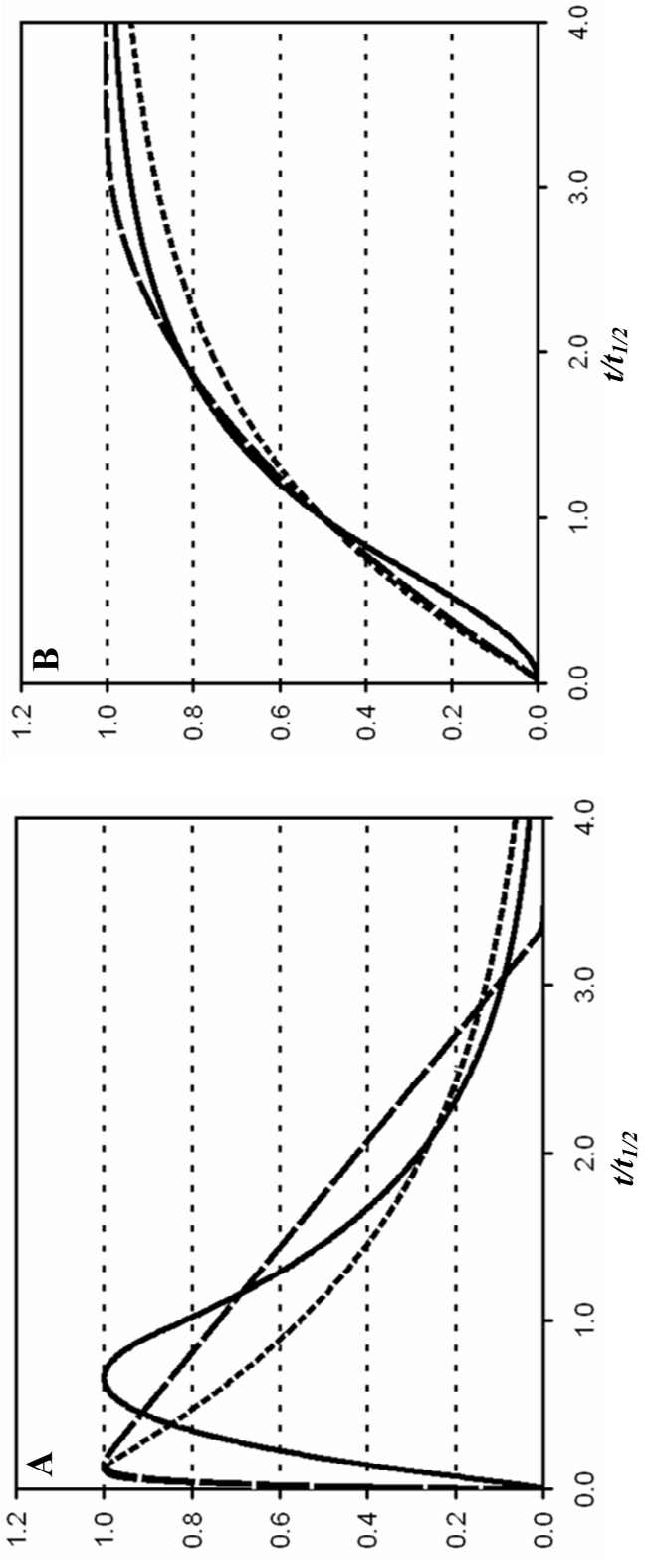
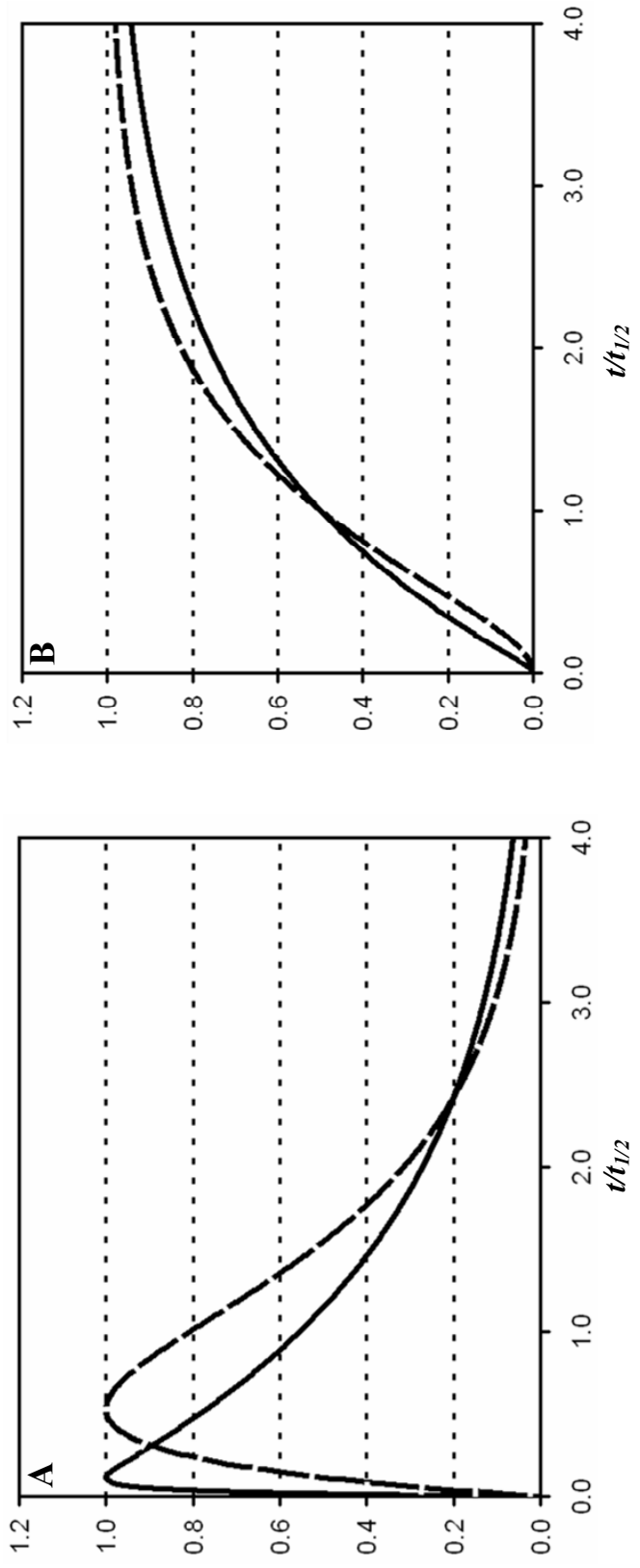

A
B
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0.0
1.0
2.0
3.0
4.0
t/t1/2

Figure 4.4: Homogenous Limit. Mixing-cup bound enzyme (Panel A, $B(t)/B_{\max}$) and product (Panel B, $P(t)/P_{\max}$) concentrations for nearly homogenous fiber hydrolysis with two different values of the binding to hydrolysis timescale ratio, ε: $\varepsilon = 0.02$ (solid), $\varepsilon = 0.2$ (dashed). The values of time (x-axis) are scaled relative to $t_{1/2}$, the value of time required to convert half of the substrate to product. In all cases $\mu = 0.5$, $\delta = 0.001$, and $\xi = 0.1$.


A
B
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0.0
1.0
2.0
3.0
4.0
t/t1/2

timescale as the fiber is consumed (Figure 4.3, Panel A). There are differences in some small-amplitude details, but even these cannot be interpreted unambiguously. For example, the product curve for concurrent hydrolysis (Figure 4.3, Panel B) shows a short-time inflection point that appears to distinguish it from the homogenous and shrinking-core cases (same figure), but an apparently identical short-time inflection point can arise in the product curve for the homogeneous case if one simply increases the ratio of binding to hydrolysis times from $\varepsilon = 0.2$ vs. $\varepsilon = 0.02$ (Figure 4.4, Panel B). Similarly, the bound enzyme curve for the shrinking-core case appears to be distinguishable from the product curves for the other two cases because it drops to zero linearly rather than decaying asymptotically, but this difference in curve-shape would disappear if there were a non-uniform distribution of fiber diameters, hence penetration times.

4.5 Discussion

We have presented a distributed binding-site model for cellulose fiber hydrolysis by cellulase enzymes, limiting our analysis to slender, dense, cylindrical fibers with uniform microstructure, in a dilute suspension, reacting at high enzyme concentrations such that the liquid phase concentration of enzyme is only slightly perturbed by the binding reaction. Moreover, we have forgone the typical "cocktail" of mechanisms included in most cellulase reaction models such as product inhibition, enzyme deactivation, synergism, irreversible binding, and so forth, as these would complicate the analysis and obfuscate the penetration-vs.-reaction scenario we are focusing on here.

We have also neglected the physical mechanisms of fiber swelling or fragmentation, which could occur as the result of hydrolysis. Interestingly, both

mechanisms would only alter how we present the radial concentration profiles. Swelling would essentially stretch the radial coordinate in the governing equations but leave the solutions unchanged, and radial fragmentation (i.e. parts of the fiber breaking away after partial reaction) would reduce the fiber size, but leave the effective reacting zone used for our calculation intact, since every point in every disconnected fragment would continue to hydrolyze at the same rate as it would if the fragment were still attached.

A comparison of the relative timescales for enzyme diffusion and hydrolysis reveals that gradients in enzyme concentration within the reacting zone are unlikely, and that the reactive zone will expand toward the center of the fiber at constant speed. To improvise for the lack of a model for how the microstructure of cellulose changes during hydrolysis, we treat this speed as an independent parameter and investigate its effect on the nature of the concentration profiles. The resulting analysis and findings differ only in the sense that the speed, although assuredly constant, is not calculated from first principles. For "slow" or "fast" penetration (small or large v), relatively large or small fractional conversions of cellulose , respectively, are required to change the unknown porosity function at the boundary's interface such that the diffusivity of the enzyme there is drastically increased ($\phi << 1$).

We identify three characteristic timescales, θ_b, $\theta_{\bar{\sigma}}$, and θ_R, associated with each of the three fundamental mechanistic processes - binding, hydrolysis, and penetration, respectively. For this analysis we assume $\theta_b << \theta_{\bar{\sigma}}$ (i.e. $\varepsilon = \frac{\theta_b}{\theta_{\bar{\sigma}}} << 1$), which leaves the relative penetration timescale, θ_R, unbounded. Thus, the relative size of θ_R, quantified by $\delta = \frac{\theta_{\bar{\sigma}}}{\theta_R}$, uniquely defines the time-evolution of the reacting zone. Concurrent penetration and hydrolysis for $\delta = 1$ is bracketed by two cases at opposite limits - homogenous for $\delta << 1$ and shrinking core for $\delta >> 1$. The shapes of the resulting radial concentration profiles are intuitive (Figure 4.2): in the homogenous

limit, horizontal lines denote the uniformity of reaction rate over radial position; in the shrinking core limit vertical lines denote the abrupt change separating as-yet-unexposed binding sites from those that are almost completely hydrolyzed; and in concurrent penetration-hydrolysis, the profile is mixed, as one would expect.

The mixing-cup concentration curves have the same general shape, regardless of which underlying mechanism (hydrolysis or penetration) limits the overall rate of consumption of the fiber or how the radial-concentration profiles evolve. The product curve rises monotonically from zero, turns over, and goes gradually toward its final maximum, and the bound enzyme curve rises to an initial maximum on a timescale that is short compared to hydrolysis, and then decays gradually to zero. This result has alarming consequences if one considers that the $B(t)$ and $P(t)$ curves are often the only time-dependent hydrolysis measurements considered by or available to researchers who study the kinetics of cellulose hydrolysis with the goal of identifying reaction mechanisms.

First consider the utility of these curves in identifying the underlying mechanism, i.e. the shape of the underlying radial concentration profiles. Admittedly, the theoretical $B(t)$ and $P(t)$ curves we generate here are not strictly identical. The product curve for concurrent hydrolysis contains a distinct short-time inflection point, which differentiates it from the two limiting cases where only simple growth is observed (Figure 4.3, Panel B). Nonetheless, product curves demonstrating this behavior should not be taken as strict indicators of concurrent hydrolysis and penetration, as the same short-time inflection can be found in product curves in the homogenous limit if the hydrolysis timescale begins to approach the timescale for enzyme binding (Figure 4.4, Panel B). The bound enzyme curve for the shrinking core limit decays linearly from its maximum, whereas in every other case there is an asymptotic decay toward zero. This special case can be attributed to our idealized

cylindrical geometry and uniform-initial binding site distribution, where the relationship between the radius and rate-of-change of volume is linear. The influence of variations in initial enzyme concentration (data not shown) on the product formation curves, assuming the parameter limitations for this model are maintained ($\psi << 1,\ \varepsilon << 1,\ \xi << 1$), offers a similarly host of ambiguities. One must conclude that the penetration mechanism and the time-dependent morphological changes in the substrate can not be identified from the shape of the enzyme binding or product curves alone.

Similarly, the $B(t)$ and $P(t)$ curves generated here would be fit well by any number of semi-mechanistic, pseudo-homogenous, Michaelis-Menten-style kinetics models available in the literature. These models contain reaction mechanisms similar to our equations (1a-b) but apply them to homogenous chemical species, generating homogenous rather than distributed, point-wise reaction rates like in our model. Subsequently, binding-site concentrations and reaction rates generated from these models only relate to the true, underlying point-wise values in special limiting circumstances. This concern also applies to so-called cellulase activities (product/time/enzyme) calculated directly from mixing-cup product curves. For example, in the tight binding limit ($\mu \to 1$) the maximum in enzyme binding relates to the total mixing-cup binding site concentration if rapid penetration occurs ($\delta << 1$), and to the surface binding site concentration if slow penetration occurs ($\delta >> 1$). Similarly, the rise-time for bound enzyme curves is related to the point-wise enzyme-binding kinetics only in two restrictive, and unlikely circumstances – either rapid (homogenous limit) or slow (shrinking core limit) penetration. The rise-time in the product curves only evidences point-wise hydrolysis kinetics in the homogenous limit. If the penetration time is long compared to the point-wise hydrolysis time, the penetration time controls the shape of the product curve. In all other circumstances in

between these limits, the maximum in enzyme binding and the rise-time for bound enzyme and product formation results from a competition of penetration and hydrolysis and are not amenable to any simple interpretation.

We conclude from this analysis that one cannot use mixing-cup product and/or enzyme-binding concentration curves alone to reliably infer how the particle morphology is changing as hydrolysis proceeds, or to determine the magnitudes of the point-wise rate constants for binding, release, and hydrolysis. One requires simultaneous measurements of the extent of hydrolysis as a function of time and penetration depth to determine if a limiting circumstance can be exploited to interpret the data, or if a detailed mechanistic model for penetration must be used.

4.6 Conclusions

We use a minimally-complex physical and mechanistic description of heterogeneous cellulose hydrolysis to develop a distributed binding-site model for hydrolysis kinetics. It includes perfect uniform cylindrical fibers, a straightforward binding and hydrolysis mechanism, and simple point-wise kinetics, yet even in this simple circumstance the shapes of the commonly measured mixing-cup enzyme binding and product curves are unreliable indicators of the underlying penetration mechanism (Table 4.1). The mixing-cup curves we generate with our penetration model could no doubt be fit with many traditional pseudo-homogeneous models for enzyme kinetics, but such fits would be purely empirical and would reveal nothing about the underlying penetration mechanism. The problem of misinterpreting data would certainly be exacerbated if the data comes from penetrating-hydrolysis of real cellulose particles with mixed geometries and physical properties, containing a non-

uniform solid phase and microstructure, and involving more complex mechanisms (product inhibition, synergism, and so forth).

This calculation reveals three important facts of use to those who study cellulose hydrolysis. First, pseudo-homogenous models for enzymatic hydrolysis, which constitute the majority of models available in the literature, are empirical, parametric curve fits that reveal little if anything about the underlying mechanism or intrinsic reaction rates. Second cellulose kinetics studies should include when possible more careful investigations of cellulose microstructure during hydrolysis, either to justify limiting penetration scenarios or to build a more reliable model for how the cellulose pore structure evolves during hydrolysis. Third, the identification or synthesis of cellulose substrates with simple, uniform geometries would improve our ability to realistically model microstructural changes during hydrolysis and understand their mechanistic implications.

4.7 Acknowledgements

This project was supported by the Initiative for Future Agriculture and Food Systems Grant no. 2001-52104-11484 from the USDA Cooperative State Research, Education, and Extension Service.

4.8 Appendix: The $O(\xi^1)$ Solution

The point-wise solutions for $\bar{z}_1$ and x_1 to equations (11a-b) at $O(\xi)$ are:

$$\bar{z}_1(T<\delta)=C_1e^{\lambda_1T'}+C_2e^{\lambda_2T'}+\cdots$$

(A1a)
$$+\frac{2}{\omega^2\delta^2}\left\{\frac{1}{2}T'^2-(1+a_7+\delta\rho)(T'-a_7)-a_2e^{\lambda_2T}+a_1a_5e^{\lambda_1T}\right\}a_3e^{\lambda_1T'}+\cdots$$
$$+\frac{2}{\omega^2\delta^2}\left\{\frac{1}{2}T'^2-(1-a_7+\delta\rho)(T'+a_7)-a_1e^{\lambda_1T}+a_2a_6e^{\lambda_2T}\right\}a_4e^{\lambda_2T'}$$

$$\bar{z}_1(T\geq\delta)=C_3e^{\lambda_1T'}+C_4e^{\lambda_2T'}+\cdots$$

(A1b)
$$-\frac{2}{\omega^2\delta^2}\left[a_3a_5e^{\lambda_1T'}-a_4e^{\lambda_2T'}\right]\left[\frac{1}{\lambda_1^2}e^{\lambda_1(T-\delta)}-a_1e^{\lambda_1T}\right]+\cdots$$
$$-\frac{2}{\omega^2\delta^2}\left[a_4a_6e^{\lambda_2T'}-a_3e^{\lambda_1T'}\right]\left[\frac{1}{\lambda_2^2}e^{\lambda_2(T-\delta)}-a_2e^{\lambda_2T}\right]$$

$$x_1(T<\delta)=-\lambda_1C_1e^{\lambda_1T'}-\lambda_2C_2e^{\lambda_2T'}+\cdots$$

(A2a)
$$+\frac{2}{\omega^2\delta^2}\left\{\begin{array}{l}-\frac{1}{2}\lambda_1T'^2-T'+(1+a_7+\delta\rho)[1+\lambda_1(T'-a_7)]+\cdots\\-\lambda_1\lambda_2a_2e^{\lambda_2T}-2a_1a_5e^{\lambda_1T}\end{array}\right\}a_3e^{\lambda_1T'}+\cdots$$
$$+\frac{2}{\omega^2\delta^2}\left\{\begin{array}{l}-\frac{1}{2}\lambda_2T'^2-T'+(1-a_7+\delta\rho)[1+\lambda_2(T'+a_7)]+\cdots\\-\lambda_1\lambda_2a_1e^{\lambda_1T}-2a_2a_6e^{\lambda_2T}\end{array}\right\}a_4e^{\lambda_2T'}$$

$$x_1(T\geq\delta)=-\lambda_1C_3e^{\lambda_1T'}-\lambda_2C_4e^{\lambda_2T'}+\cdots$$

(A2b)
$$+\frac{2}{\omega^2\delta^2}\left[2\lambda_1a_3a_5e^{\lambda_1T'}+\lambda_1\lambda_2a_4e^{\lambda_2T'}\right]\left[\frac{1}{\lambda_1^2}e^{\lambda_1(T-\delta)}-a_1e^{\lambda_1T}\right]+\cdots$$
$$+\frac{2}{\omega^2\delta^2}\left[2\lambda_2a_4a_6e^{\lambda_2T'}+\lambda_1\lambda_2a_3e^{\lambda_1T'}\right]\left[\frac{1}{\lambda_2^2}e^{\lambda_2(T-\delta)}-a_2e^{\lambda_2T}\right]$$

The constant terms in these solutions are:

$$C_1 = \frac{2}{\delta^2\omega^2}\left[-(1-\mu)a_1 e^{\lambda_1\delta(1-\rho)} + a_7(\lambda_1 a_3 - \lambda_2 a_4 a_6)a_2 e^{\lambda_2\delta(1-\rho)} + a_7(1-a_7+\delta\rho)a_4\right]$$

$$C_2 = \frac{2}{\delta^2\omega^2}\left[-(1-\mu)a_2 e^{\lambda_2\delta(1-\rho)} - a_7(\lambda_2 a_4 - \lambda_1 a_3 a_5)a_1 e^{\lambda_1\delta(1-\rho)} - a_7(1+a_7+\delta\rho)a_3\right]$$

$$C_3 = C_1 - \frac{2}{\omega^2\delta^2}\left[\left(1-\varepsilon+\delta\rho+\frac{1}{2}\delta^2\rho^2\right)a_3 - a_4 a_6 a_7^2 e^{(\lambda_2-\lambda_1)\delta\rho}\right]$$

$$C_4 = C_2 - \frac{2}{\omega^2\delta^2}\left[\left(1-\varepsilon+\delta\rho+\frac{1}{2}\delta^2\rho^2\right)a_4 - a_3 a_5 a_7^2 e^{(\lambda_1-\lambda_2)\delta\rho}\right]$$

The lumped parameters in these solutions are:

$$a_1 = \frac{1-\lambda_1\delta}{\lambda_1^2} \quad a_2 = \frac{1-\lambda_2\delta}{\lambda_2^2} \quad a_3 = \frac{1+\mu\lambda_1}{\lambda_1} \quad a_4 = \frac{1+\mu\lambda_2}{\lambda_2}$$

$$a_5 = \frac{1}{1+2\omega\lambda_1} \quad a_6 = \frac{1}{1-2\omega\lambda_2} \quad a_7 = \frac{1}{\lambda_1-\lambda_2}$$

The mixing-cup solutions for $\overline{Z}_1$ and X_1 from equations (9a-b) at $O(\xi)$ are:

(A3a) $$\overline{Z}_1(T<\delta) = \frac{4}{\omega^2\delta^4}\left\{\begin{matrix} A_0 + A_1 T + A_2 T^2 + A_3 e^{\lambda_1 T} + A_4 e^{\lambda_2 T} + A_5 T e^{\lambda_1 T} + A_6 T e^{\lambda_2 T} + \cdots \\ + A_7 e^{(\lambda_1+\lambda_2)T} + A_8 T^2 e^{\lambda_1 T} + A_9 T^2 e^{\lambda_2 T} + A_{10} e^{2\lambda_1 T} + A_{11} e^{2\lambda_2 T} \end{matrix}\right\}$$

(A3b) $$\overline{Z}_1(T \geq \delta) = \frac{4}{\omega^2 \delta^4} \left\{ \begin{array}{l} A_{12} e^{\lambda_1 T} + A_{13} e^{\lambda_2 T} + A_7 e^{(\lambda_1+\lambda_2)T} + A_{10} e^{2\lambda_1 T} + \cdots \\ + A_{11} e^{2\lambda_2 T} + A_{14} e^{\lambda_1 (T-\delta)} + A_{15} e^{\lambda_2 (T-\delta)} + A_{16} e^{2\lambda_1 (T-\delta)} + \cdots \\ + A_{17} e^{2\lambda_2 (T-\delta)} + A_{18} e^{(\lambda_1+\lambda_2)(T-\delta)} + A_{19} e^{\lambda_1 T} e^{\lambda_1 (T-\delta)} + \cdots \\ + A_{20} e^{\lambda_2 T} e^{\lambda_2 (T-\delta)} + A_{21} e^{\lambda_1 (T-\delta)} e^{\lambda_2 T} + A_{22} e^{\lambda_1 T} e^{\lambda_2 (T-\delta)} + \cdots \\ + A_{23} e^{\lambda_1 \delta} e^{\lambda_2 (T-\delta)} + A_{24} e^{\lambda_2 \delta} e^{\lambda_1 (T-\delta)} \end{array} \right\}$$

(A4a) $$X_1(T < \delta) = \frac{4}{\omega^2 \delta^4} \left\{ \begin{array}{l} - A_1 - 2A_2 T + A_{25} e^{\lambda_1 T} + A_{26} e^{\lambda_2 T} + A_{27} T e^{\lambda_1 T} + \cdots \\ + A_{28} T e^{\lambda_2 T} - \lambda_1 A_8 T^2 e^{\lambda_1 T} - \lambda_2 A_9 T^2 e^{\lambda_2 T} + \cdots \\ + \lambda_1 \lambda_2 A_7 e^{(\lambda_1+\lambda_2)T} - 2\lambda_1 A_{10} e^{2\lambda_1 T} - 2\lambda_2 A_{11} e^{2\lambda_2 T} \end{array} \right\}$$

(A4b) $$X_1(T \geq \delta) = \frac{4}{\omega^2 \delta^4} \left\{ \begin{array}{l} - \lambda_1 A_{12} e^{\lambda_1 T} - \lambda_2 A_{13} e^{\lambda_2 T} + \lambda_1 \lambda_2 A_7 e^{(\lambda_1+\lambda_2)T} + \cdots \\ - 2\lambda_1 A_{10} e^{2\lambda_1 T} - 2\lambda_2 A_{11} e^{2\lambda_2 T} - \lambda_1 A_{14} e^{\lambda_1 (T-\delta)} + \cdots \\ - \lambda_2 A_{15} e^{\lambda_2 (T-\delta)} - 2\lambda_1 A_{16} e^{2\lambda_1 (T-\delta)} - 2\lambda_2 A_{17} e^{2\lambda_2 (T-\delta)} + \cdots \\ + \lambda_1 \lambda_2 A_{18} e^{(\lambda_1+\lambda_2)(T-\delta)} - 2\lambda_1 A_{19} e^{\lambda_1 T} e^{\lambda_1 (T-\delta)} + \cdots \\ - 2\lambda_2 A_{20} e^{\lambda_2 T} e^{\lambda_2 (T-\delta)} + \lambda_1 \lambda_2 A_{21} e^{\lambda_1 (T-\delta)} e^{\lambda_2 T} + \cdots \\ + \lambda_1 \lambda_2 A_{22} e^{\lambda_1 T} e^{\lambda_2 (T-\delta)} - \lambda_2 A_{23} e^{\lambda_1 \delta} e^{\lambda_2 (T-\delta)} + \cdots \\ - \lambda_1 A_{24} e^{\lambda_2 \delta} e^{\lambda_1 (T-\delta)} \end{array} \right\}$$

The lumped parameters in these solutions are:

$$A_0 = \left\{ (1-\mu) \left[1 - 2\varepsilon + (2+\delta)^2 \right] - \varepsilon (2+\delta) \right\} \omega^2$$

$$A_1 = \left[\varepsilon - 2(1-\mu)(2+\delta) \right] \omega^2$$

$$A_2 = (1-\mu) \omega^2$$

$$A_3 = a_7^3 B_1 + a_7^3 B_2 - a_1 a_2 a_4 + a_1^2 a_3 + \frac{1}{\lambda_1^2} a_1^2 a_3 a_5 + \frac{1}{\lambda_1^4}(1-\lambda_1) a_3 + \cdots$$
$$+ \frac{1}{\lambda_1^2}(1-\lambda_1) a_1 a_3 - (1-a_7) a_1 a_4 a_7 + \left(\frac{1}{\lambda_1^3} + \lambda_1 a_1^2\right) a_4 a_7 - (1+\delta+a_7) a_1 a_3 a_7$$

$$A_4 = -a_7^3 B_1 - a_7^3 B_2 - a_1 a_2 a_3 + a_2^2 a_4 + \frac{1}{\lambda_2^2} a_2^2 a_4 a_6 + \frac{1}{\lambda_2^4}(1-\lambda_2) a_4 + \cdots$$
$$+ \frac{1}{\lambda_2^2}(1-\lambda_2) a_2 a_4 + (1+a_7) a_2 a_3 a_7 - \left(\frac{1}{\lambda_2^3} + \lambda_2 a_2^2\right) a_3 a_7 + (1+\delta-a_7) a_2 a_4 a_7$$

$$A_5 = \frac{1}{\lambda_1}\left(\frac{1}{\lambda_1^2} - a_7^2\right) a_1 a_3 a_5 - \left(\frac{1}{\lambda_2^2} - a_7^2\right) \lambda_2 a_1 a_4 + (1+\delta+a_7) a_1 a_3 - (1-\mu)\delta a_1$$

$$A_6 = \frac{1}{\lambda_2}\left(\frac{1}{\lambda_2^2} - a_7^2\right) a_2 a_4 a_6 - \left(\frac{1}{\lambda_1^2} - a_7^2\right) \lambda_1 a_2 a_3 + (1+\delta-a_7) a_2 a_4 - (1-\mu)\delta a_2$$

$$A_7 = a_1 a_2 (a_3 + a_4)$$

$$A_8 = \frac{1}{2}(1-\mu-a_3) a_1$$

$$A_9 = \frac{1}{2}(1-\mu-a_4) a_2$$

$$A_{10} = -\frac{1}{\lambda_1^2} a_1^2 a_3 a_5$$

$$A_{11} = -\frac{1}{\lambda_2^2} a_2^2 a_4 a_6$$

$$A_{12} = a_7^3 B_1 + a_1 B_3 - a_1 B_5 - \frac{1}{2}\delta^2(1-\mu-a_3) a_1 + \frac{1}{\lambda_1^2}\delta^2(1-\lambda_1) a_3 + \frac{1}{\lambda_1}\delta^2 a_4 a_7$$

$$A_{13} = -a_7^3 B_2 + a_2 B_6 + a_2 B_4 - \frac{1}{2}\delta^2(1-\mu-a_4) a_2 + \frac{1}{\lambda_2^2}\delta^2(1-\lambda_2) a_4 - \frac{1}{\lambda_2}\delta^2 a_3 a_7$$

$$A_{14} = -\frac{1}{\lambda_1^2} B_3 + \frac{1}{\lambda_1^2} B_5 + \frac{1}{\lambda_2^4} a_4 a_6 a_7^2$$

$$A_{15} = -\frac{1}{\lambda_2^2} B_4 - \frac{1}{\lambda_2^2} B_6 + \frac{1}{\lambda_1^4} a_3 a_5 a_7^2$$

$$A_{16} = -\frac{1}{\lambda_1^6} a_3 a_5$$

$$A_{17} = -\frac{1}{\lambda_2^6} a_4 a_6$$

$$A_{18} = \frac{1}{\lambda_1^2 \lambda_2^2} (a_3 + a_4)$$

$$A_{19} = 2\frac{1}{\lambda_1^4} a_1 a_3 a_5$$

$$A_{20} = 2\frac{1}{\lambda_2^4} a_2 a_4 a_6$$

$$A_{21} = -\frac{1}{\lambda_1^2} (a_3 + a_4) a_2$$

$$A_{22} = -\frac{1}{\lambda_2^2} (a_3 + a_4) a_1$$

$$A_{23} = -\frac{1}{\lambda_1} \left(\frac{1}{\lambda_1} - a_7 \right) a_1 a_3 a_5 a_7^2 - \lambda_2 a_1 a_4 a_7^3$$

$$A_{24} = -\frac{1}{\lambda_2} \left(\frac{1}{\lambda_2} + a_7 \right) a_2 a_4 a_6 a_7^2 + \lambda_1 a_2 a_3 a_7^3$$

$$A_{25} = -\lambda_1 a_7^3 B_1 - \lambda_2 a_7^3 B_2 - 2\frac{1}{\lambda_1} a_1^2 a_3 a_5 - \lambda_1 \lambda_2 a_1 a_2 a_4 - (a_3 + a_4) \lambda_1^2 a_1^2 a_7 + \cdots$$

$$+ \lambda_1 \left(a_7 + a_7^2 - \frac{2}{\lambda_1^2} \right) a_1 a_3 - \frac{1}{\lambda_1^3} (1 - \lambda_1) a_3 + (1 - a_7) \lambda_1 a_1 a_4 a_7 - \frac{1}{\lambda_1^2} a_4 a_7$$

$$A_{26} = \lambda_1 a_7^3 B_1 + \lambda_2 a_7^3 B_2 - 2a_2^2 a_4 \frac{1}{\lambda_2} a_6 - \lambda_1 \lambda_2 a_1 a_2 a_3 - a_7 (1 + a_7) \lambda_2 a_2 a_3 + \cdots$$
$$+ 2a_2 a_3 a_7 + \delta^2 a_3 a_7 + \lambda_2^2 a_2^2 a_4 a_7 - \lambda_2 \left(a_7 - a_7^2 + \frac{2}{\lambda_2^2} \right) a_2 a_4 - \frac{1}{\lambda_2^3} (1 - \lambda_2) a_4$$

$$A_{27} = -\left(\lambda_1 + \lambda_2^2 a_7^2\right) a_1 a_4 + \frac{1}{\lambda_1} \left(-\frac{2}{\lambda_1} + \lambda_2 a_7^2 \right) a_1 a_3 a_5 + \cdots$$
$$+ \lambda_1^2 a_1^2 a_3 - (1 + a_7) \lambda_1 a_1 a_3 + (1 - \mu) \delta \lambda_1 a_1$$

$$A_{28} = -\left(\lambda_2 + \lambda_1^2 a_7^2\right) a_2 a_3 + \frac{1}{\lambda_2} \left(-\frac{2}{\lambda_2} + \lambda_1 a_7^2 \right) a_2 a_4 a_6 + \cdots$$
$$+ \lambda_2^2 a_2^2 a_4 - (1 - a_7) \lambda_2 a_2 a_4 + (1 - \mu) \delta \lambda_2 a_2$$

The additional lumped parameters are:

$$B_1 = \left(\frac{1}{\lambda_2} a_4 a_6 - \lambda_1 a_3 \right) (1 - \lambda_1 \delta + \lambda_2 \delta) a_2$$

$$B_2 = \left(\frac{1}{\lambda_1} a_3 a_5 - \lambda_2 a_4 \right) (1 + \lambda_1 \delta - \lambda_2 \delta) a_1$$

$$B_3 = \left(1 - \frac{1}{\lambda_1 \lambda_2} - \frac{2}{\lambda_1} + \frac{3}{\lambda_1^2} \right) a_3$$

$$B_4 = \left(1 - \frac{1}{\lambda_1 \lambda_2} - \frac{2}{\lambda_2} + \frac{3}{\lambda_2^2} \right) a_4$$

$$B_5 = \left(1 - a_7 - \frac{2}{\lambda_1} \right) a_4 a_7$$

$$B_6 = \left(1 + a_7 - \frac{2}{\lambda_2} \right) a_3 a_7$$

The mixing-cup solution for $Y(T)$ to equation (7c) is:

(A5a) $Y(T < \delta) = 1 - \left(1 - \frac{T}{\delta}\right)^2 - \overline{Z}(T < \delta)$;

(A5b) $Y(T \geq \delta) = 1 - \overline{Z}(T \geq \delta)$.

APPENDIX A

HYDROLYSIS OF ELECTROSPUN CELLULOSE FIBERS*

*Reproduced in part with permission from Park HJ, Dingee JW, Fitzgibbon SR, Anton AB, and Joo YL (2007) *J. Biobased Materials and Bioenergy* **1** (2), 1-12 © 2007 American Scientific Publishers.

The following study involves an electrospun cellulose substrate which was produced and provided by Heidi J. Park (Cornell University, Chemical Engineering MS 2009), in the lab of Dr. Yong L. Joo, School of Chemical and Biomolecular Engineering, Cornell University.

A.1 Introduction

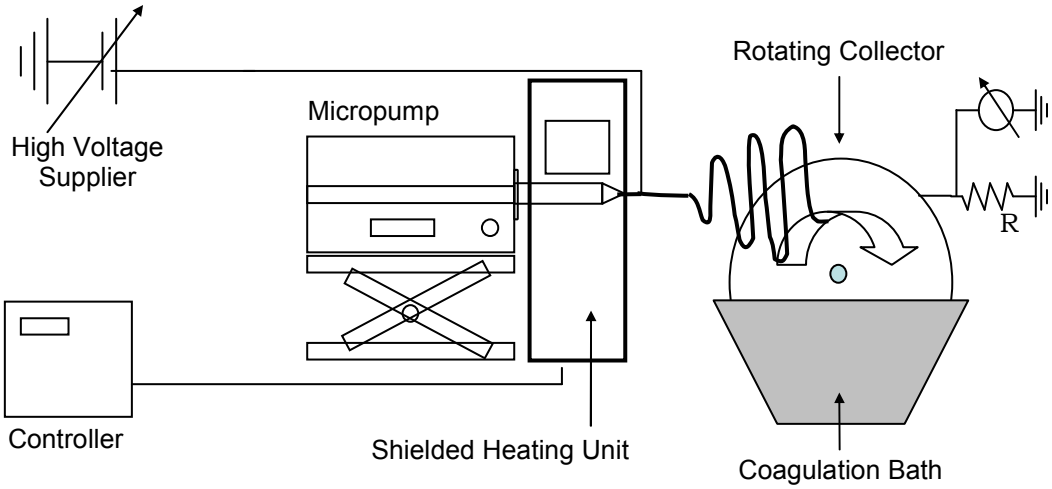
The enzymatic depolymerization of cellulose to soluble sugars offers a potential feedstock for the production renewable commodities and fuel **[1]**. Cellulose occurs naturally as a heterogeneous, insoluble mixture of crystalline and amorphous phases, often embedded in a matrix with other plant-cell-wall carbohydrates and proteins **[16]**. Relatively pure "model" cellulose substrates are created by physical/chemical treatments that remove impurities and reducing particle size **[10]**. However, even within these model substrates, the complex multiphase, insoluble microstructure of cellulose remains intact. Moreover, each model substrate is characterized a set of different physical properties, which makes it difficult to systematically study the relationship between individual properties and hydrolyzability. A microscopically uniform, insoluble cellulose substrate with well

defined and *controllable* physical properties could offer a distinct advantage over current model substrates. Such a substrate may be available through the process of electrostatic fiber spinning, which produces stable non-woven cellulose fibers **[165, 166]**.

Electrostatic fiber spinning or 'electrospinning' is a novel process for forming fibers with submicron diameters through the action of electrostatic forces. Here, a polymer (in our case cellulose) solution or melt is charged by placing it in a capillary and connecting it to a high voltage supply, as shown in Figure A.1. When the electrical force at the interface of a polymer solution overcomes the surface tension, a charged jet is ejected. The jet initially extends in a straight line then undergoes a vigorous whipping motion caused by electrohydrodynamic instability **[167].** As the solvent evaporates, the polymer is collected on a grounded mesh or plate in the form of non-woven fibers of high surface-area-to-mass ratio. Fibers produced by this electrostatic spinning process can be several orders of magnitude smaller than those produced by conventional fiber-spinning methods **[168, 169]**. The size and physical properties of the cellulose fiber is controlled by source of cellulose used in the melt, the solvent system used, and the processing conditions. Electrospinning allows one to manufacture cellulose fibers with dimensions comparable to natural biomass fibers, but with variable DP, crystallinity, and fiber diameter.

Herein we investigate the binding and hydrolysis of electrospun cellulose (ESC) fibers by *Thermobifida fusca* Cel5A, a retaining endo-specific cellulase enzyme. The binding and hydrolysis kinetics are similar to other insoluble model substrates, with cellobiose being the dominant soluble product. However, the morphological changes caused by Cel5A hydrolysis are counter-intuitive. SEM imaging of hydrolyzed fibers reveals fiber slicing with loss of long range connectivity rather than fiber thinning. Although the mechanisms for these morphological changes

Figure A.1: Electrospinning Apparatus. The electrospinning apparatus consists of a grounded metal collector (mesh or plate) and a syringe positioned horizontally with its needle at high electric potential. The cellulose solution (1 to 9 wt%) is fed mechanically with a micropump, and the voltage is provided by a high-voltage power supply. A motor attached to the collector rotates the collector through the coagulant water bath.


Rotating Collector
Micropump
High Voltage
Supplier
R
Controller
Shielded Heating Unit
Coagulation Bath

are unresolved, ESC remains an interesting substrate for cellulase mechanistic studies.

A.2 Methods

A.2.1 Enzyme Production & Purification

S. lividans strain PGG74 expressing Cel5A was grown as described by Jung *et al.* **[60]** in a 10 liter culture for 2-3 days at dissolved oxygen levels no less than 30%, impeller speeds of 300-320 rpm, pH of 7.0 and 7.5, and aeration rates of 12 l/min. The production culture was harvested by centrifugation at 4,000 rpm for 30 min at 5 ^{o}C. The supernatant was collected and poured over glass wool, and ammonium sulfate (AS) and PMSF were added to 1.2 M and 0.1 mM, respectively. The supernatant was then clarified by centrifugation at 8,000 rpm for 30 min at 5 ^{o}C. This supernatant was collected and clarified by depth microfitration using a 2 μm Cuno BetaPure polyolefin cartridge, item number AU09Z13NG020 (Cuno Inc, Meriden CT).

The filtrate was further purified at 5 ^{o}C as described by Irwin *et al.* **[137]** using a P-Sepharose hydrophobicity column followed by a Q-Sepharose anionexchange column with the following modifications. The loaded P-Sepharose column was washed with 1-column volume of 1.2 M AS, 2-column volumes of 0.6 M AS, 10 mM NaCl, and 5 mM Kpi, pH 6.0, followed by 3-column volumes of 0.3 M AS, 5 mM NaCl, and 5 mM Kpi, pH 6.0, and eluted with 3-column volumes of 5 mM Kpi, pH 6.0. The loaded Q-Sepharose column was eluted with a 6-column volume gradient of 0-0.3M NaCl, 1-10 mmho, in 10mMBisTris of pH 5.5, followed by a 2-column volume wash with 0.5 M NaCl and 10 mM BisTris, 20 mmho, pH 5.5.

Purified Cel5A was exchanged into standard buffer (50 mM NaAc and 0.02% NaAz, pH 5.5) and concentrated to roughly 60 μM using a stirred cell ultrafiltration

chamber with a polyethersulfone 10 kDa MWCO ultrafiltration membrane (Millipore, Billerica, MA). All purified cellulase concentrations were determined spectrophometrically with published extinction coefficients **[170]**. The resulting protein was stored at -70 ^{o}C.

A.2.2 ESC Production

Cellulose was electrospun from solutions of cellulose/N-methylmorpholine-N-oxide (NMMO)/water as described by Park **[171]**. Degree of polymerization (DP) 210 cellulose (Whatman CF-11 powder) was dissolved in 85/15 NMMO/water (w/w) at concentrations of 9 wt%.

ESC was characterized before and after hydrolysis for fiber diameter, crystallinity, and crystal structure. Fiber diameter and morphology was observed via scanning electron microscope (LEICA 440 SEM). Wide angle X-ray scattering (WAXS, Scintag, Inc. Theta-Theta Diffractometer) was used to determine the crystal structure and degree of crystallinity of the samples. The degree of crystallinity was obtained by taking the area ratio of the crystalline phase to the sum of the crystalline plus amorphous phases, which was obtained after deconvolution of each peak in the WAXS patterns **[165]**.

A.2.3 Binding

Cel5A was bound to electrospun cellulose in 2-ml screw-cap centrifuge tubes and rotated end-over-end at 5 ^{o}C with 10 mg/ml ESC or 1 mg/ml BMCC and varying amounts of Cel5A in standard buffer. All materials were pre-chilled for 24 hrs at 5 ^{o}C, and all reactions were carried out in triplicate. Duplicate enzyme and buffer blanks

and single substrate blanks were made for each concentration. After binding for 2 hrs, the tubes were centrifuged at 5 ^{o}C and 13,000 rpm for 10 min to form pellets. The supernatant was removed from the pellet, centrifuged again, and a second supernatant collected. These supernatants were analyzed for protein concentration by A280 using the extinction coefficient for Cel5A, 97,100 M^{-1}.

A.2.4 Hydrolysis

ESC was hydrolyzed at 50 ^{o}C in 2-ml screw-cap centrifuge tubes rotated end-over-end, each with a total volume of 1.6 ml containing 4 mg/ml ESC and 1 μM Cel5A in standard buffer. Substrate was pre-suspended in buffer for 24 hrs prior to hydrolysis, and all reactions were carried out in triplicate. Duplicate enzyme and buffer blanks and single substrate blanks were made for each time point. At each time point, triplicate sample tubes along with their respective blanks were centrifuged at room temperature at 13,000 rpm for 5 min to form pellets. The supernatant was removed, centrifuged again, and a second supernatant was then collected. The second supernatant was then analyzed for reducing end concentration using the PAHBAH assay and for sugar composition using the TLC assay, both described by Irwin *et al.* **[137]**. The remaining pellets were resuspended in distilled water, vortexed, and centrifuged again at 13,000 rpm for 5 min. This washing procedure was repeated 3 times. The wet pellets were frozen with dry ice and lyophilized for 3 days before being analyzed with SEM.

A.3 Results

Cel5A bound to ESC during hydrolysis, observed as a reduction in the free

enzyme concentration in the liquid-phase. Binding was characterized at 5 ^{o}C, where the hydrolysis process is slowed considerably. For comparison, isotherms were constructed for both ESC and BMCC (Figure A.2).

ESC was hydrolyzed by Cel5A, with conversion reaching nearly 80% after more than 200 hours of exposure at 50 ^{o}C (Figure A.3, Left Panel). The dominant reducing sugar product was cellobiose (Figure A.3, Right Panel).

SEM analysis of hydrolyzed fibers revealed fiber deterioration at 20% conversion with increasing deterioration approaching 100% conversion (Figure A.4).

A.4 Discussion

ESC offers a novel model-insoluble substrate for use in mechanistic hydrolysis studies, with the distinct benefit of having systematically controllable physical properties **[171]**. Here we investigate the ability of a single-purified cellulase enzyme to hydrolyze ESC spun from an NMMO/water melt with a low DP (~210). Regeneration of cellulose after direct dissolution results in dramatic changes in both the cellulose crystal structure and microstructure. Moreover, the crystal structure the ESC used here was Type II, which is different from the usual Type I found in naturally occurring cellulose.

Cellulase binding is an important aspect of the hydrolysis process because it positions the enzymes catalytic active site at the liquid-solid phase boundary. Cel5A bound to ESC in much the same way it binds to other insoluble substrates. We characterized binding with a low-temperature isotherm at 5 ^{o}C, where the hydrolysis reaction is slowed considerably (Figure A.2). Although such isotherms are not transferable to the high-temperature reaction at 50 ^{o}C, they do indicate the extent of binding that can occur on the substrate's surface. On a per mass basis, Cel5A was

Figure A.2: Low-Temperature Binding Isotherm. Bound vs. free enzyme for Cel5A binding at 5 ^{o}C on BMCC (top isotherm, left vertical axis) and ESC (bottom isotherm, right vertical axis). ESC used in this assay had an average fiber diameter of ~0.5 μm and an average crystallinity of ~50%.

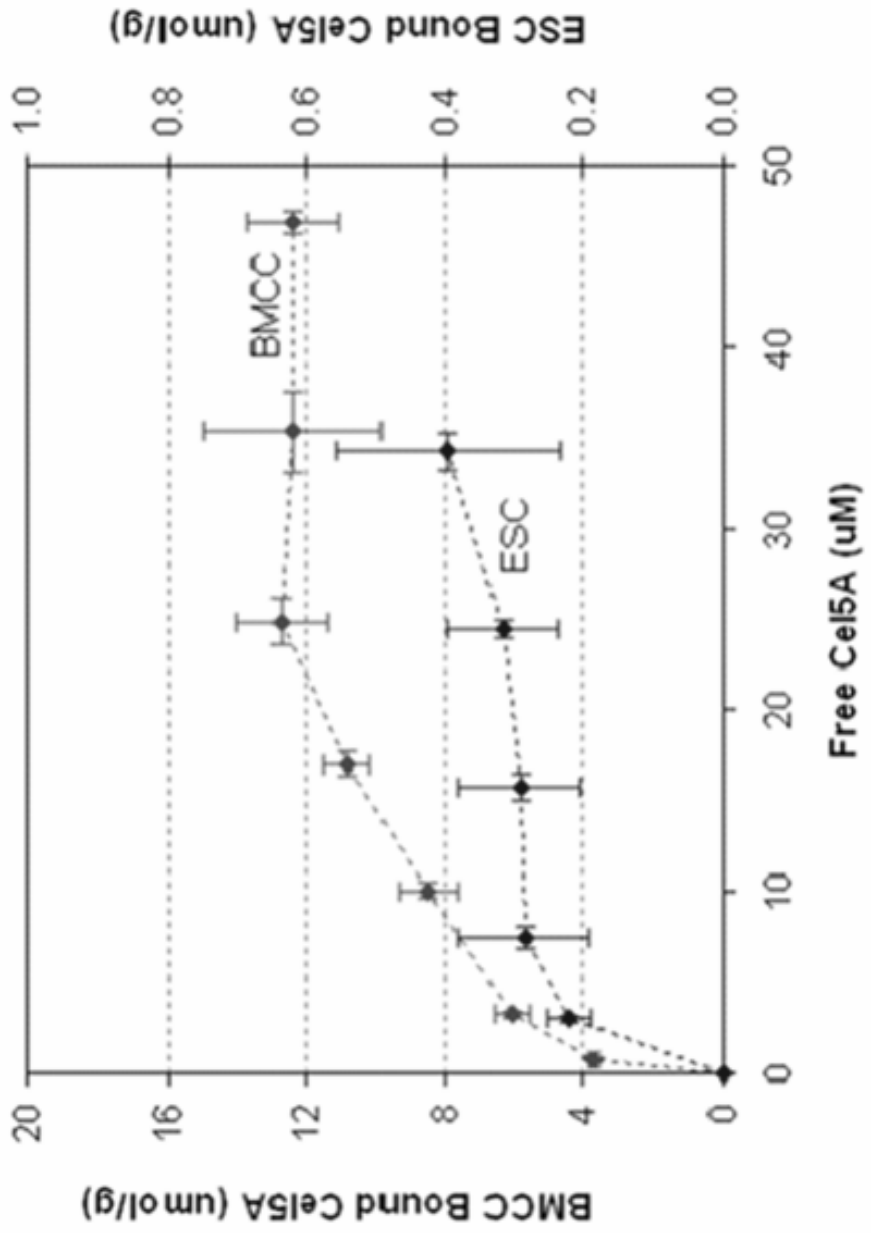
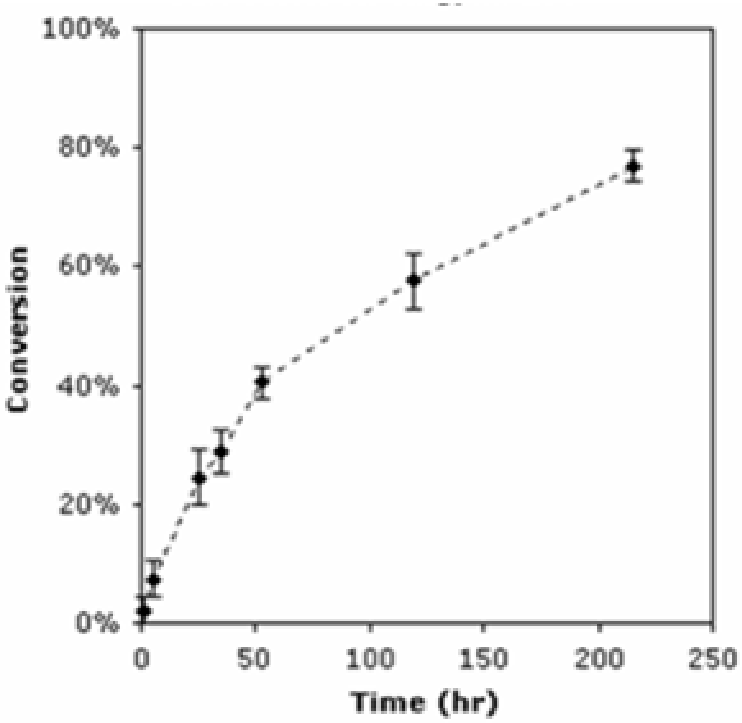

ESC Bound Cel5A (umol/g)
1.0
0.8
0.6
0.4
0.2
0.0
BMCC
ESC
0
10
20
30
40
50
Free Cel5A (uM)
BMCC Bound Cel5A (umol/g)
20
16
12
8
4
0

Figure A.3: ESC Hydrolysis with Cel5A. Conversion of ESC (4mg/ml) by Cel5A (1 μM) at 50 ^{o}C (left) presented with TLC plate for reducing sugar compositions (right). Conversion calculated from reducing sugar concentrations as determined by PAHBAH assay, using an approximate 1:8:1 ratio of G1:G2:G3 reducing sugar products. TLC plate is read top-to-bottom with the first column containing a bench (B) with G1-G4 and subsequent columns indicating sugar composition for time-points shown in the conversion plot. The intensities of the bands approximate the relative mass of sugars present. ESC used in this assay had an average fiber diameter of ~1.0 μm and average crystallinity of ~62%.


100%
80%
60%
40%
20%
0%
Conversion
0
50
100
150
200
250
Time (hr)

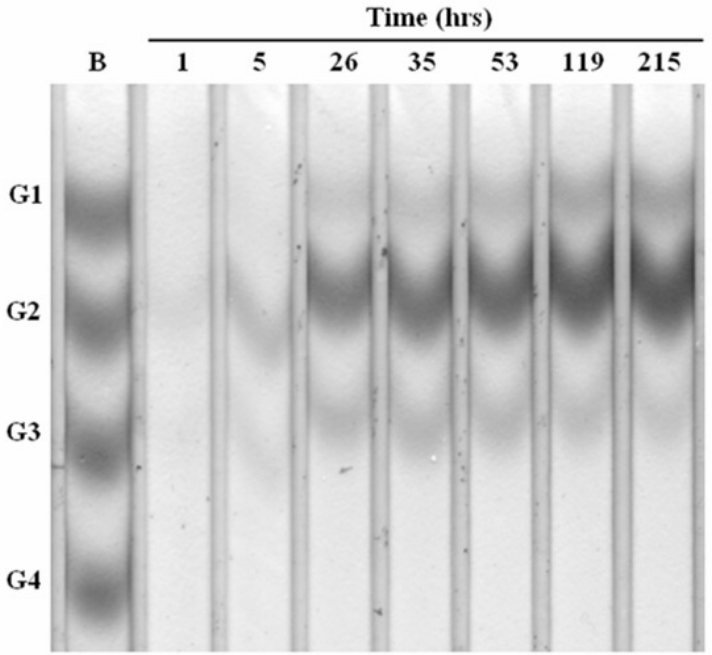
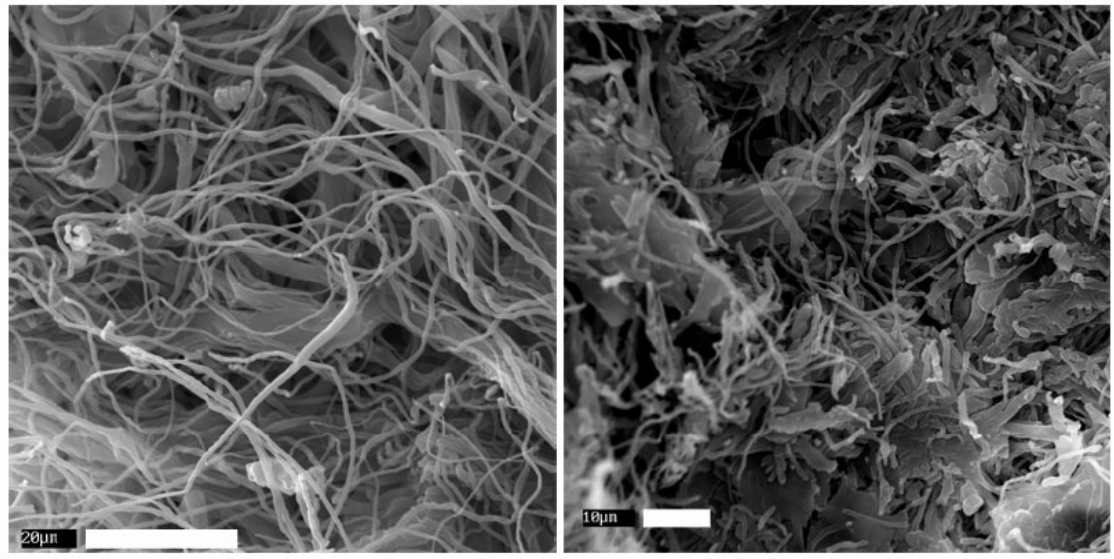
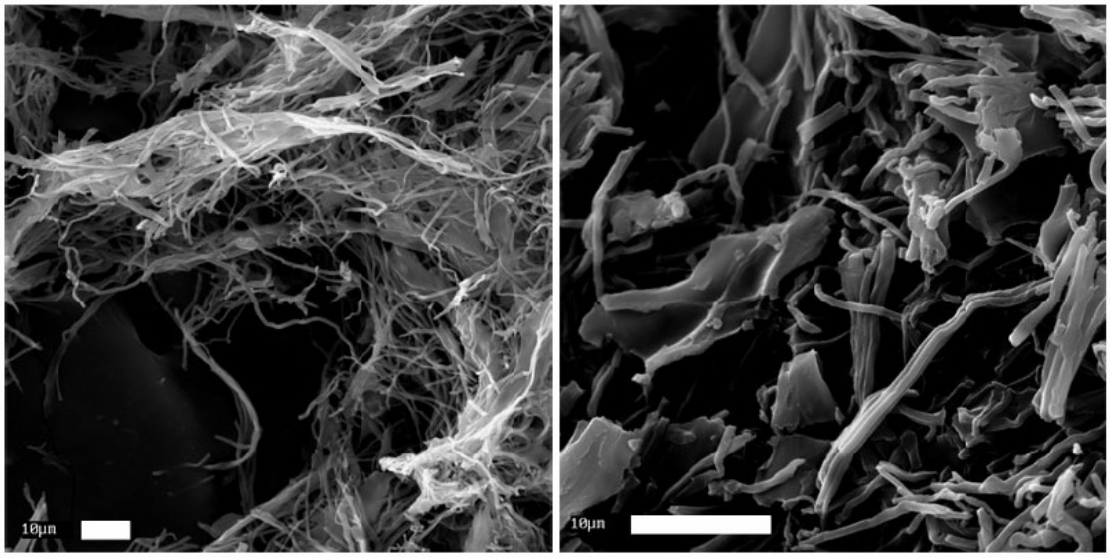

Time (hrs)
B
1
5
26
35
53
119
215
G1
G2
G3
G4

Figure A.4: SEM Images of ESC after Exposure to Cel5A. Exposure time and conversion for each panel as follows: (a) 1 hr or 2% conversion; (b) 35 hrs or 29% conversion; (c) 53 hrs or 40% conversion; (d) 215 hrs or 77% conversion. Hydrolysis data corresponding to these images is shown in Figure A.3.



(a) (b)



(c) (d)

able to bind to BMCC (12 μmol/g) much more than ESC (0.3 μmol/g). The different loading capacities can be explained by the difference in specific surface areas of the two substrates: ESC (5.5 m^2/g) vs. BMCC (200 m^2/g) **[75]**.

The electrospinning process drastically alters the morphology and crystallinity of the cellulose substrate, which may in turn alter its reactivity. Nonetheless, Cel5A was able to hydrolyze ESC, converting more the half the substrate to reducing sugars over timescales typically observed for other model substrates (Figure A.3, Left Panel). Moreover, the reducing sugar composition, observed by TLC, showed a dominant cellobiose product, with trace concentrations of glucose (G1) and cellotriose (G3). These product compositions are consistent hydrolysis data for Cel5A reported previously by Barr **[138]**.

ESC's uniform cylindrical microstructure offers a distinct advantage over other model substrates. SEM images of ESC used in this study reveal a relatively well behaved microstructural system of non-woven cylindrical fibers. Structural abnormalities including fiber matting, parallel fiber aggregation, and large fiber formation, were observed but they pale in comparison to the wide range of structural heterogeneities and mixed characteristic dimensions observed for comparative model substrates like BMCC. Moreover, these abnormalities may be removed with improvements in the electropinning process. SEM analysis of ESC after exposure to Cel5A showed fiber degradation, starting at 20% conversion and increasing toward 100% conversion (Figure A.4). The predominant fiber degradation mechanism appeared to be fiber slicing and fragmentation with loss of long-range fiber connectivity. Moreover, at 40% conversion a low-density precipitate began to appear with the original fiber fragments after washing and lyophilization, and became the dominant recovered solid phase approaching 100% conversion. SEM images of the secondary precipitate revealed fibers with characteristic lengths of 10 μm, with

increased surface charging caused possibly by a roughening of the surface by the cellulase enzyme. The reasons for the formation of these fragments remain unclear. They may be recalcitrant artifacts of the electrospinning process, freed by the enzyme hydrolysis reaction. Alternatively, mechanically weaker sections of fiber that have been attacked by the cellulase enzymes could break apart during the post-hydrolysis recover process. The later explanation may explain the apparent lack of fiber thinning, which one might expect to observe prior to fragmentation.

A.5 Conclusions

ESC offers a novel alternative to conventional model cellulose substrates, where fiber morphology and physical properties can be controlled by the choice of starting material, solvent used, and various processing conditions **[171]**. Stable-non-woven cellulose fibers ranging in diameter from 0.1 μm to 10 μm, containing up to 80% crystallinity, and DPs between 200 and 1000 have been successfully created **[165, 166, 171]**. Application of ESC in cellulase studies is limited by production rate, reproducibility, and fiber uniformity, all of which could be corrected through improvements in process control. ESC fibers contain Type II crystal structure, which may play an important role in the hydrolysis process.

The feasibility of ESC as a model substrate has been demonstrated. Fibers bind cellulase enzymes and are hydrolyzed in much the same way as other model substrates, such as BMCC, Avicel, and filter paper. SEM reveals the morphology of ESC changes over the course of hydrolysis, showing fragmentation and loss of long-range connectivity. These morphological changes are contrary to expected fiber thinning, which was not observed but would be expected for dense-cylindrical fibers

being hydrolyzed from their outer surfaces. Hydrolysis also results in the formation of short-length fiber fragments with characteristic lengths around 10 μm.

ESC remains an interesting alternative to model substrates but process improvements must be developed to ensure more concise reproducibility and higher substrate yields. Moreover, the mechanism for morphological level changes caused by hydrolysis must be better understood, possibly thought the use of coaxial fibers containing an outer shell of cellulose with a nonhydrolyzable fiber core **[171]**.

A.6 Acknowledgments

The ESC used in this study was provided by Yong L. Joo's research group at Cornell University. The author would like to personally thank Heidi J. Park for her collaborative work in all aspects of this project and for the substantial quantities of ESC she produced for it.

REFERENCES

1. Wyman CE. Potential synergies and challenges in refining cellulosic biomass to fuels, chemicals, and power. Biotechnol Prog 2003;19(2):254-62.
2. Wyman CE. What is (and is not) vital to advancing cellulosic ethanol. Trends Biotechnol 2007;25(4):153-7.
3. Krässig HA. Cellulose: Structure, Accessibility, and Reactivity. Yverdon, Switzerland: Gordon and Breach Science Publishers; 1993.
4. Nevell T, Zeronian SH, eds. Cellulose Chemistry and Its Applications. New York: John Wiley & Sons; 1985.
5. Teleman A, Koivula A, Reinikainen T, et al. Progress-curve analysis shows that glucose inhibits the cellotriose hydrolysis catalyzed by cellobiohydrolase II from *Trichoderma reesei*. Eur J Biochem 1995;231(1):250-8.
6. Gardner K, Blackwell J. The hydrogen bonding in cellulose. Biochim Biophys Acta 1974;343:232-7.
7. Gardner K, Blackwell J. The structure of native cellulose. Biopolymers 1974;13:1975–2001.
8. O'Sullivan AC. Cellulose: The structure slowly unravels. Cellulose 1997;4(3):173-207.
9. Atalla RH, VanderHart DL. Native cellulose: a composite of two distinct crystalline forms. Science 1984;223(4633):283-5.
10. Mansfield SD, Mooney CA, Saddler JN. Substrate and enzyme characteristics that limit cellulose hydrolysis. Biotechnol Prog 1999;15:804-16.
11. Klemm D, Heublein B, Fink H-P, Bohn A. Cellulose: Fascinating biopolymer and sustainable raw material. Angew Chem Int Ed 2005; 44:3358-93.

12. Kadla JF, Gilbert RD. Cellulose structure: A review. Cell Chem Technol 2000;34(3-4):197-216.
13. Mohr H, Schopfer P. Plant Physiology. Berlin: Springer-Verlag; 1995.
14. Mohnen D. Biosynthesis of pectins and galactomannans. In: Pinto BM, ed. Comprehensive Natural Products Chemistry. Amsterdam: Elsevier; 1999:497-527.
15. Somerville C, Bauer S, Brininstool G, et al. Toward a systems approach to understanding plant cell walls. Science 2004;306:2206 - 11.
16. Rose JKC, ed. Annual Plant Reviews. Oxford: Wiley-Blackwell; 2003.
17. Hoshino E, Nomura M, Takai M, Okazaki M, Nisizawa K, Kanda T. Action of exo- and endo-type cellulases from *Irpex lacteus* on Valonia cellulose. J Ferment Bioeng 1994;77(5):496-502.
18. Blackwell J, Kolpak FJ. The cellulose microfibril as an imperfect array of elementary fibrils. Macromolecules 1975;8(3):322-6.
19. Wood TM. Preparation of crystalline, amorphous, and dyed cellulase substrates. Method Enzymol 1988;160:19-25.
20. Kuhad RC, Singh A, Eriksson KE. Microorganisms and enzymes involved in the degradation of plant fiber cell walls. Adv Biochem Engin/Biotechnol 1997;57:45-125.
21. Coutinho P, Henrissat B. The Molecular Structure of Cellulases and Other Carbohydrate-Active Enzymes: An Integrated Database Approach. In: Ohmiya K, Hayashi K, Sakka K, Kobayahsi Y, Karita S, Kimura T, eds. Genetics, Biochemistry and Ecology of Cellulose Degradation. Tokyo, Japan: Uni Publishers Co.; 1999:15-23.
22. Wilson DB, Irwin DC. Genetics and properties of cellulases. Adv Biochem Eng 1999;65:1-21.

23. Mandels M, Reese ET. Induction of cellulase in *Trichoderma viride* as influenced by carbon sources and metals. J Bacteriol 1957;73(2):269-78.

24. Mandels M, Weber J, Parizek R. Enhanced cellulase production by a mutant of *Trichoderma viride*. Appl Microbiol 1971;21(1):152.

25. Schülein M. Enzymatic properties of cellulases from *Humicola insolens*. J Biotechnol 1997;57:71-81.

26. Wilson DB. Studies of *Thermobifida fusca* plant cell wall degrading enzymes. Chem Rec 2004;4:72-82.

27. Bray MR, Creagh AL, Damude HG, et al. β-1,4-glycanases of *Cellulomonas fimi*: Families, Mechanisms, and Kinetics. In: Enzymes for Pulp and Paper Processing. Washington: AMER CHEMICAL SOC; 1996:64-84.

28. Lamed R, Setter E, Kenig R, Bayer EA. The cellulosome - A discrete cell-surface organelle of *Clostridium thermocellum* which exhibits separate antigenic, cellulose-binding and various cellulolytic activities. Biotechnol Bioeng 1983;163-181.

29. Desvauxa M. The cellulosome of *Clostridium cellulolyticum*. Enzyme Microb Technol 2005;37:373-85.

30. Desvauxa M. *Clostridium cellulolyticum*: Model organism of mesophilic cellulolytic clostridia. FEMS Microbiol Rev 2005;29(4):741-64.

31. Bayer EA, Chanzy H, Lamed R, Shoham Y. Cellulose, cellulases, and cellulosomes. Curr Opin Struct Biol 1998;8(5):548-57.

32. Cohen R, Suzuki MR, Hammel KE. Processive endoglucanase active in crystalline cellulose hydrolysis by the brown rot basidiomycete *Gloeophyllum trabeum*. Appl Environ Microbiol 2005;71(5):2412-7.

33. Li YC, Irwin DC, Wilson DB. Processivity, substrate binding, and mechanism of cellulose hydrolysis by *Thermobifida fusca* Ce19A. Appl Environ Microbiol 2007;73:3165-72.

34. Boisset C, Fraschini C, Schülein M, Henrissat B, Chanzy H. Imaging the enzymatic digestion of bacterial cellulose ribbons revals the endo character of the cellobiohydrolase Cel6A from *Humicola insolens* and its mode of synergy with cellobiohydrolase Cel7A. Appl Environ Microbiol 2000;66:1444-52.

35. Teeri TT. Crystalline cellulose degradation: New insight into the function of cellobiohydrolases. Trends Biotechnol 1997;15(5):160-7.

36. Henrissat B. A classification of glycosyl hydrolases bases on amino acid sequence similarities. Biochem J 1991;280:309-16.

37. Henrissat B, Bairoch A. Updating the sequence-based classification of glycosyl hydrolases. Biochem J 1996;316:695-6.

38. Boraston AB, Bolam DN, Gilbert HJ, Davies GJ. Carbohydrate-binding modules: Fine-tuning polysaccharide recognition. Biochem Eng J 2004;382:769-81.

39. Davies GJ. Structure studies on cellulases. Biochem Soc Trans 1998;26:167-73.

40. Rabinovich ML, Melnick MS, Bolobova AV. The structure and mechanism of action of cellulolytic enzymes. Biochemistry (Moscow) 2002;67(8):850-71.

41. Henrissat B, Teeri TT, Warren RAJ. A scheme for designating enzymes that hydrolyse the polysaccharides in the cell walls of plants. FEBS Lett 1998;425:352-4.

42. Koshland PJ. Stereochemistry and the mechanism of enzymatic reactions. Biol Rev 1953;28:416-36.

43. Sinnott ML. Catalytic mechanisms of enzymic glycosyl transfer. Chem Rev 1990;90:1171-202.

44. Lodge JA, Maier T, Liebl W, Hoffmann V, Strater N. Crystal structure of *Thermotoga maritima* α-glucosidase AglA defines a new clan of NAD+-dependent glycosidases. J Biol Chem 2003;278:19151-8.

45. Rajan SS, Yang X, Collart F, et al. NAD-dependent hydrolysis by family 4 glycosidases involves a novel elimination mechanism. Structure 2004;12:1619-29.

46. Varrot A, Yip VLY, Li Y, et al. NAD+ and metal-ion dependent hydrolysis by family 4 glycosidases: structural insight into specificity for phospho-β-D-glucosides. J Mol Biol 2005;346:423-35.

47. Gebler J, Gilkes NR, Claeyssens M, et al. Stereoselective hydrolysis catalyzed by related β-1,4-glucanases and β-1,4-xylanases. J Biol Chem 1992;267(18):12559-61.

48. Boraston AB, McLean BW, Kormos JM, et al. Carbohydrate-Binding Modules: Diversity of Structure and Function. In: Gilbert HJ, Davies GJ, Henrissat B, Svensson B, eds. Recent Advances in Carbohydrate Bioengineering. Cambridge: Royal Society of Chemistry; 1999:202–11.

49. Bray MR, Johnson PE, Gilkes NR, McIntosh LP, Kilburn DG, Warren RAJ. Probing the role of tryptophan residues in a cellulose-binding domain by chemical modification. Protein Science 1996;5(11):2311-8.

50. Tomme P, Driver DP, Amandoron EA, et al. Comparison of a Fungal (Family-I) and Bacterial (Family-Ii) Cellulose-Binding Domain. J Bacteriol 1995;177(15):4356-63.

51. Wilson DB, Spezio M, Irwin DC, Karplus A, Taylor J. Comparison of enzymes catalyzing the hydrolysis of insoluble polysaccharides. ACS Symp Ser 1995;618:1-12.

52. Din N, Gilkes NR, Tekant B, Miller RC, Warren AJ, Kilburn DG. Non-hydrolytic disruption of cellulose fibers by the binding domain of a bacterial cellulase. Bio-Technology 1991;9(11):1096-9.

53. Banka RR, Mishra S. Adsorption properties of the fibril forming protein from *Trichoderma reesei*. Enzyme Microb Technol 2002;31:784-93.

54. Banka RR, Mishra S, Ghose TK. Fibril formation from cellulose by a novel protein from *Trichoderma reesei*: A non-hydrolytic cellulolytic component? World J Microbiol and Biotechnol 1998;14(4):551-8.

55. Harrison MJ, Nouwens AS, Jardine DR, et al. Modified glycosylation of cellobiohydrolase I from high cellulase-producing mutant strain of *Trichoderma reesei*. Eur J Biochem 1998;256:119-27.

56. Receveur W, Czjzek M, Schülein M, Panine P, Henrissat B. Dimension, shape, and conformational flexibility of a two domain fungal cellulase in solution probed by small angle X-ray scattering. J Biol Chem 2002;277(43):40887-92.

57. Srisodsuk M, Reinikainen T, Penttilä M, Teeri TT. Role of the interdomain linker peptide of *Trichoderma reesei* cellobiohydrolase I in its interaction with crystalline cellulose. J Biol Chem 1993;268(28):20756-61.

58. Chandra RP, Bura R, Mabee WE, Berlin A, Pan X, Saddler JN. Substrate pretreatment: The key to effective enzymatic hydrolysis of lignocellulosics? Adv Biochem Engin/Biotechnol 2007;108:67-93.

59. Donner TR, Evans BR, Affholter KA, Woodward J. Role of cellulose-binding domain of cellobiohydrolase I in cellulose hydrolysis. ACS Symp Ser 1994;566:75-83.

60. Jung H, Wilson DB, Walker LP. Binding and reversibility of *Thermobifida fusca* Cel5A, Cel6B, and Cel48A and their respective catalytic domains to bacterial microcrystalline cellulose. Biotechnol Bioeng 2003;84:151-9.

61. Nieves RA, Ellis RP, Todd RJ, Johnson TJA, Grohmann K, Himmel ME. Visualization of *Trichoderma reesei* cellobiohydrolase I and endoglucanase I on aspen cellulose by using monoclonal anti-body gold conjugates. Appl Environ Microbiol 1991;57(11):3163-70.

62. Lehtiö J, Sugiyama J, Gustavsson M, Fransson L, Linder M, Teeri TT. The binding specificity and affinity determinants of family 1 and family 3 cellulose binding modules. PNAS 2003;100(2):484-9.

63. Jervis EJ, Haynes CA, Kilburn DG. Surface diffusion of cellulases and their isolated binding domains on cellulose. J Biol Chem 1997;272(38):24016-23.

64. Hildén L, Johansson G. Recent developments on cellulases and carbohydrate-binding modules with cellulose affinity. Biotechnol Lett 2004;26:1683-93.

65. Palonen H, Tenkanen M, Linder M. Dynamic interaction of *Trichoderma reesei* cellobiohydrolases Ce16A and Ce17A and cellulose at equilibrium and during hydrolysis. Appl Environ Microbiol 1999;65(12):5229-33.

66. Yang B, Willies DM, Wyman CE. Changes in the enzymatic hydrolysis rate of avicel cellulose with conversion. Biotechnol Bioeng 2006;94(6):1122-8.

67. Nidetzky B, Steiner W, Claeyssens M. Cellulose hydrolysis by the cellulases from *Trichoderma reesei*: Adsorptions of 2 cellobiohydrolases, 2 endocellulases and their core proteins on filter-paper and their relation to hydrolysis. Biochem J 1994;303:817-23.

68. Bothwell MK, Wilson DB, Irwin DC, Walker LP. Binding reversibility and surface exchange of *Thermomonospora fusca* E_3 and E_5 and *Trichoderma reesei* CBHI. Enzyme Microb Technol 1997;20:411-7.

69. Linder M, Teeri TT. The cellulose-binding domain of the major cellobiohydrolase of *Trichoderma reesei* exhibits true reversibility and a high exchange rate on crystalline cellulose. PNAS 1996;93(22):12251-5.

70. Lee I, Evans BR, Woodward J. The mechanism of cellulase action on cotton fibres: Evidence from atomic force microscopy. Ultramicroscopy 2000;82:213-21.

71. Cao Y, Tan H. Structural characterization of cellulose with enzymatic treatment. J Mol Struct 2004;705:189-93.

72. Gama FM, Carvalho MG, Figueiredo MM, Mota M. Comparative study of cellulose fragmentation by enzymes and ultrasound. Enzyme Microb Technol 1997;20:12-7.

73. Walker LP, Wilson DB, Irwin DC. Measureing fragmentation of cellulose by *Thermomonospora fusca* cellulase. Enzyme Microb Technol 1990;12:378-86.

74. Jung H, Wilson DB, Walker LP. Binding of *Thermobifida fusca* CD_{Cel5A}, CD_{Cel6B}, and CD_{Cel48A} to easily hydrolysable and recalcitrant cellulose fractions on BMCC. Enzyme Microb Technol 2002;31:941-8.

75. Zhang Y-HP, Lynd LR. Toward an aggregated understanding of enzymatic hydrolysis of cellulose: Noncomplexed cellulase systems. Biotechnol Bioeng 2004;88(7):797-824.

76. Fan LT, Gharpuray MM, Lee YH. Cellulose Hydrolysis. Berlin: Springer-Verlag; 1987.

77. Holtzapple MT, Caram HS, Humphrey AE. Determining the inhibition constants in the HCH-1 model of cellulose hydrolysis. Biotechnol Bioeng 1984;26(7):753-7.

78. Holtzapple MT, Caram HS, Humphrey AE. The HCH-1 model of enzymatic cellulose hydrolysis. Biotechnol Bioeng 1984;26(7):775-80.

79. Gan Q, Allen SG, Taylor G. Kinetic dynamics in heterogeneous enzymatic hydrolysis of cellulose: An overview, an experimental study and mathematical modelling. Process Biochem 2003;38(7):1003-18.

80. Nidetzky B, Steiner W, Hayn M, Claeyssens M. Cellulose hydrolysis by the cellulases from *Trichoderma reesei*: A new model for synergistic interaction. Biochem J 1994;298:705-10.

81. South CR, Hogsett DAL, Lynd LR. Modeling simultaneous saccharification and fermentation of lignocellulose to ethanol in batch and continuous reactors. Enzyme Microb Technol 1995;17(9):797-803.

82. Zhang Y-HP, Lynd LR. A functionally based model for hydrolysis of cellulose by fungal cellulase. Biotechnol Bioeng 2006;94(5):888-.

83. Converse AO, Optekar JD. A synergistic kinetics model for enzymatic cellulose hydrolysis compared to degree-of-synergism experimental results. Biotechnol Bioeng 1993;42(1):145-8.

84. Suga K, van Dedem G, Moo-Young M. Degradation of polysaccharides by endo and exo enzymes: A theoretical analysis. Biotechnol Bioeng 1975;17:433-9.

85. Okazaki M, Moo-Young M. Kinetics of enzymatic hydrolysis of cellulose: Analytical description of a mechanistic model. Biotechnol Bioeng 1978;20:637-63.

86. Fenske JJ, Penner MH, Bolte JP. A simple individual-based model of insoluble polysaccharide hydrolysis: the potential for autosynergism with dual-activity glycosidases. J Theor Biol 1999;199:113-8.

87. Väljamäe P. The Kinetics of Cellulose Enzymatic Hydrolysis: Implications of the Synergism Between Enzymes [PhD]. Uppsala: University of Uppsala; 2002.

88. Väljamäe P, Sild V, Pettersson LG, Johansson G. The initial kinetics of hydrolysis by cellobiohydrolases I and II is consistent with a cellulose surface-erosion model. Eur J Biochem 1998;253:469-75.

89. Sild V. Modeling of Enzyme Kinetics on Polymeric Substrates by Simulation [PhD]. Uppsala: Uppsala University; 1999.

90. Sild V, Ståhlberg J, Pettersson LG, Johansson G. Effect of potential binding site overlap to binding of cellulase to cellulse: a two-dimensional simulation. FEBS Lett 1996;378:51-6.

91. Lin CC, Segel LA. Mathematics Applied to Deterministic Problems in the Physical Sciences. New York: SIAM; 1988.

92. Claeyssens M, Aerts G. Characterization of cellulolytic activities in commercial *Trichoderma reesei* preparations: An approach using small, chromogenic substrates. Bioresource Technol 1992;39(2):143-6.

93. van Tilbeurgh H, Claeyssens M, de Bruyne CK. The use of 4-methylumbelliferyl and other chromophoric glycosides in the study of celluloytic enzymes. FEBS Lett 1982;149(1):152-6.

94. van Tilbeurgh H, Loontiens FG, de Bruyne CK, Claeyssens M. Fluorogenic and chromogenic glycosides as substrates and ligands of carbohydrases. Methods Enzymol 1988;160:45-59.

95. Garrett RH, Grisham CM. Biochemistry. 3rd ed. Belmont, CA: Thomson Brooks/Cole; 2005.

96. Nelson DL, Cox MM. Lehninger Principles of Biochemistry. 4th ed. New York: Worth Publishing; 2000.

97. Shuler ML, Kargi F. Bioprocess Engineering: Basic Concepts. 2nd ed. Upper Saddle River, NJ: Prentice Hall; 2002.

98. Michaelis L, Menten ML. Die kinetik der invertinwirkung. Biochem Z 1913;49:333-69.

99. Hatzimanikatis V, Li CH, Ionita JA, Henry CS, Jankowski MD, Broadbelt LJ. Exploring the diversity of complex metabolic networks. Bioinformatics 2005;21:1603-9.

100. Wolkenhauer O, Mesarovic M. Feedback dynamics and cell function: Why systems biology is called systems biology. Mol BioSystems 2005;1(1):14-6.

101. Van Slyke D, Cullen GE. The mode of action of urease and enzymes in general. J Biol Chem 1914;19:141-80.

102. Bender CM, Orszag SA. Advanced Mathematical Methods for Scientists and Engineers: Asymptotic Methods and Perturbation Theory. New York, NY: Springer; 1999.

103. Holmes MH. Introduction to Perturbation Methods. New York: Springer-Verlag; 1995.

104. Dean WM. Analysis of Transport Phenomena. New York: Oxford University Press; 1998.

105. Leal LG. Advanced Transport Phenomena: Fluid Mechanics and Convective Transport Processes. New York: Cambridge University Press; 2007.

106. Aris R. The Mathematical Theory of Diffusion and Reaction in Permeable Catalysts. England: Oxford University Press; 1975.

107. Jiaqi M. The singularly perturbed problem for combustion reaction diffusion. Acta Math Appl Sin-E 2001;17(2):255-9.

108. van Harten A. Applications of Singular Perturbation Techniques to Combustion Theory. In: Theory and Applications of Singular Perturbations. Berlin: Springer; 1982.

109. Vasileva B, Stelmakh VG. Singularly distributed systems in the theory of semiconductor devices. USSR Comput Math Phys 1977;17:48-58.

110. Ardema MD. Singular Perturbations in Systems and Control. In: CISM Courses and Lectures. New York: Springer; 1983:180.

111. Kokotovic PV, Hassan KK, O'Reilly J. Singular Perturbation Methods in Control: Analysis and Design. In: SIAM Classics in Applied Mathematics; 1999:25.

112. Bodenstein MZ. Phyzik Chem 1913;85:329.

113. Bowen J, Acrivos A, Oppenheim A. Singular perturbation refinement to quasi-steady state approximation in chemical kinetics. Chem Eng Sci 1963;18:177-88.

114. Briggs GE, Haldane JBS. A note on the kinetics of enzyme action. Biochem J 1925;19(2):338-9.

115. Tzafriri AR. Michaelis-Menten kinetics at high enzyme concentrations. Bull Math Biol 2003;65:1111-29.

116. Tzafriri AR, Edelman ER. Quasi-steady-state kinetics at enzyme and substrate concentrations in excess of the Michaelis-Menten constant. J Theor Biol 2007;245:737-48.

117. Segel LA, Slemrod M. The quasi-steady-state assumption: A case study in perturbation. SIAM Review 1989;31(3):446-77.

118. Borghans JAM, de Boer RJ, Segel LA. Extending the quasi-steady state approximation by changing variables. Bull Math Biol 1996;58:43-63.

119. Schnell S, Maini PK. Enzyme kinetics at high enzyme concentrations. Bull Math Biol 2000;62:483-99.

120. Schnell S, Mendoza C. The condition for pseudo-first-order kinetics in enzymatic reactions is independent of the initial enzyme concentration. Biophys J 2004;107:165-74.

121. Segel LA. On the validity of the steady state assumption of enzyme kinetics. Bull Math Biol 1988;50:579-93.

122. Laidler KJ. Theory of the transient phase in kinetics, with special reference to enzyme systems. Can J Chem 1955;33:1614-24.

123. Seshadri MS, Fritzsch G. Analytical solutions of a simple enzyme kinetic problem by a perturbative procedure. Biophys Struct Mech 1980;6:111-23.

124. Heineken FG, Tsuchiya HM, Aris R. On the mathematical status of the pseudo-steady state hypothesis of biochemical kinetics. Math Biosci 1967;1:95-113.

125. Laidler KJ. Chemical Kinetics. 3rd ed. New York: Prentice-Hall; 1987.

126. Moore JW, Pearson RG. Kinetics and Mechanism. 3rd ed. New York: Wiley; 1981.

127. Rubinow SI. Introduction to Mathematical Biology. New York: Wiley; 1975.

128. Murray JD. Mathematical Biology: An Introduction. 3rd ed. New York: Springer; 2002.

129. Ashino R, Nagase M, Vaillancourt R. Behind and beyond the MATLAB ODE suite. Comput Math Appl 2000;40:491-512.

130. Shampine LF, Reichelt MW. The MATLAB ODE suite. J Sci Comput 1997;18(1):1-22.

131. Withers SG. Mechanisms of glycosyl transferases and hydrolases. Carbohydr Polym 2001;44(4):325-37.

132. Brumer H, Sims PFG, Sinnott ML. Lignocellulose degradation by *Phanerochaete chrysosporium*: Purification and characterization of the main α-galactosidase. Biochem J 1999;339:43-53.

133. Kawai R, Igarashi K, Kitaoka M, Ishii T, Samejima M. Kinetics of substrate transglycosylation by glycoside hydrolase family 3 glucan (1→3)-β-glucosidase from the white-rot fungus *Planerochaete chrysosporium*. Carbohydr Res 2004;339(18):2851-7.

134. Claeyssens M, van Tilbeurgh H, Kamerling JP, Berg J, Vršanká M, Biely P. Studies of the cellulolytic system of the filamentous fungus *Trichoderma reesei* QM 9414: Substrate specificity and transfer activity of endoglucanase I. Biochem J 1990;270:251-6.

135. Schmid G, Wandrey C. Characterization of a cellodextrin glucohydrolase with soluble oligomeric substrates: Experimental results and modeling of concentration-time-course data. Biotechnol Bioeng 1989;33(11):1445-60.

136. Savel'ev AN, Ibatyllin FM, Eneyskaya EV, Kachurin AM, Neustroev KN. Enzymatic properties of α-D-galactosidase from *Trichoderma reesei*. Carbohydr Res 1996;296:261-73.

137. Irwin DC, Cheng M, Xiang B, Rose JKC, Wilson DB. Cloning, expression and characterization of a family-74 xyloglucanase from *Thermobifida fusca*. Eur J Biochem 2003;270(14):3083-91.

138. Barr BK. Hydrolysis, Specificity, and Active-Site Binding of Glycosides by *Thermomonospora fusca* Cellulases [PhD]. Ithaca, NY: Cornell University; 1997.

139. Chirico WJ, Brown RD. Separation of [1-^{3}H] cellooligosaccharides by thin-layer chromatography: Assay for cellulolytic enzymes. Anal Biochem 1985;150(2):264-72.

140. Jung ED, Lao G, Irwin DC, Barr BK, Benjamin A, Wilson DB. DNA sequences and expression in *Streptomyces lividans* of an exoglucanase gene and an endoglucanase gene from *Thermomonospora fusca*. Appl Environ Microbiol 1993;59(9):3032-43.

141. Bevington PR, Robinson DK. Data Reduction and Error Analysis for the Physical Sciences. 3rd ed. Boston: McGraw-Hill; 2003.

142. Sakon J, Adney WS, Himmel ME, Thomas SR, Karplus PA. Crystal structure of thermostable family 5 endocellulase E1 from *Acidothermus cellulolyticus* in complex with cellotetraose. Biochemistry 1996;35(33):10648-60.

143. Chapon V, Czjzek M, El Hassouni M, Py B, Juy M, Barras F. Type II protein secretion in gram-negative pathogenic bacteria: The study of the structure/secretion relationships of the cellulase CeI5 (formerly EGZ) from *Erwinia chrysanthemi*. J Mol Biol 2001;310(5):1055-66.

144. Davies GJ, Dauter M, Brzozowski AM, Bjørnvad ME, Andersen KV, Schülein M. Structure of the *Bacillus agaradherans* family 5 endoglucanase at 1.6 Å and its cellobiose complex at 2.0 Å resolution. Biochemistry 1998;37:1926-32.

145. Shaw A, Bott R, Vonrhein C, Bricogne G, Power S, Day AG. A novel combination of two classic catalytic schemes. J Mol Biol 2002;320(2):303-9.

146. Davies GJ, Wilson KS, Henrissat B. Nomenclature for sugar-binding subsites in glycosyl hydrolases. Biochem J 1997;321:557-9.

147. Davies GJ, Mackenzie L, Varrot A, et al. Snapshots along an enzymatic reaction coordinate: Analysis of a retaining β-glycoside hydrolase. Biochemistry 1998;37:11707-13.

148. Harjunpää V, Teleman A, Koivula A, Ruohonen L, Teeri TT. Cello-oligosaccharide hydrolysis by cellobiohydrolase II from *Trichoderma reesei*:

Association and rate constants derived from an analysis of progress curves. Eur J Biochem 1996;240:584-91.

149. Nidetzky B, Zachariae W, Gercken G, Hayn M, Steiner W. Hydrolysis of cellooligosaccharides by *Trichoderma reesei* cellobiohydrolases: Experimental data and kinetic modeling. Enzyme Microb Technol 1994;16(1):43-52.

150. Shieh W-J, Tsai GJ, Ladisch MR, Tsao GT. Ultraviolet difference spectroscopic study on the interactions of cellulase from *Trichoderma reesei* with cellodextrins. Appl Biochem Biotechnol 1989;22(1):13-29.

151. Wood TM, McCrae SI. The purification and properties of the C_1 component of *Trichoderma koningii* cellulase. Biochem J 1972;128:1183-92.

152. Holtzapple M, Cognata M, Shu Y, Hendrickson C. Inhibition of *Trichoderma reesei* cellulase by sugars and solvents. Biotechnol Bioeng 1990;36(3):275-87.

153. Dingee JW, Anton AB. A new perturbation solution to the Michaelis-Menten problem. AIChE J 2008;54(5):1344-57.

154. Movagharnejad K, Sohrabi M, Kaghazchi T, Vahabzadeh F. A model for the rate of enzymatic hydrolysis of cellulose in heterogeneous solid-liquid systems. Biochem Eng J 2000;4(3):197-206.

155. Movagharnejad K. Modified shrinking particle model for the rate of enzymatic hydrolysis of impure cellulosic waste materials with enzyme reuse by the substrate replacement. Biochem Eng J 2005;24(33):217-23.

156. Movagharnejad K, Sohrabi M. A model for the rate of enzymatic hydrolysis of some cellulosic waste materials in heterogeneous solid-liquid systems. Biochem Eng J 2003;14:1-8.

157. Wang L, Zhang Y, Gao P, Shi D, Liu H, Gao H. Changes in the structural properties and tate of hydrolysis of cotton fibers during extended enzymatic hydrolysis. Biotechnol Bioeng 2006;93:443-56.

158. Szekely J, Evans JW. A structural model for gas-solid reactions with a moving boundary. Chem Eng Sci 1970;25:1091-107.

159. Gavalas GR. A random capillary model with application to char gasification at chemically controlled rates. AIChE J 1980;26:577-85.

160. Brem G, Brouwers JJH. Analytical solutions for non-linear conversion of a porous solid particle in a gas -- I. Isothermal conversion. Chem Eng Sci 1990;45:1905-13.

161. Deen WM. Hindered transport of large molecules in liquid-filled pores. AIChE J 1987;33:1409-25.

162. Bishop MT, Langley KH, Karasz FE. Dynamic light-scattering studiers of polymer diffusion in porous materials: Linear polystylene in porous glass. Macromolecules 1989;22:1220-31.

163. Whitaker DR, Colvin JR, Cook WH. Molecular weight and shape of *Myrothecium verrucaria* cellulase. Arch Biochem Biophys 1954;49:257-62.

164. Gavals GR. A Random Capillary Model with Application to Char Gasification at Chemically Controlled Rates. AIChE Journal 1980;26:577-85.

165. Kim CW, Kim DS, Kang SY, Marquez M, Joo YL. Structural studies of electrospun cellulose nanofibers. Polymer 2006;47:5097-107.

166. Kim CW, Frey MW, Marquez M, Joo YL. Preparation of electrospun cellulose nanofibers via direct dissolution. J Polym Sci Pol Phys 2005;43:1673-83.

167. Fong H, Reneker DH. Electrospinning and the Formation of Nanofibers. In: Salem DR, ed. Structure Formation in Polymeric Fibers. Cincinatti: Hanser Gardner Publications, Inc.; 2001:225-46.

168. Ramakrishna S, Fujihara K, Teo W-E, Lim T-C, Ma Z. An Introduction to Electrospinning and Nanofibers. Singapore: World Scientific Publishing Co, Pte. Ltd.; 2005.

169. Subbiah T, Bhat GS, Tock RW, Parameswaran S, Ramkumar SS. Electrospinning of nanofibers. J Appl Polym 2005;96(2):557-69.

170. Irwin DC, Spezio M, Walker LP, Wilson DB. Activity studies of eight purified cellulases: Specificity, synergism, and binding domain effects. Biotechnol Bioeng 1993;42(8):1002-13.

171. Park HJ. Electrospun Cellulose as a Model Substrate for Enzymatic Hydrolysis. Ithaca, NY: Cornell University; 2009.